DOCUMENT RESUME

ED 423 342 UD 032 547

AUTHOR	Luhm, Theresa; Foley, Ellen; Corcoran, Tom
TITLE	The Accountability System: Defining Responsibility for Student Achievement. Children Achieving: Philadelphia's Education Reform. Progress Report Series 1996-1997.
INSTITUTION	Consortium for Policy Research in Education, Philadelphia, PA.; Research for Action, Inc., Philadelphia, PA.; OMG Center for Collaborative Learning, Philadelphia, PA.
SPONS AGENCY	Children Achieving Challenge, Philadelphia, PA.
PUB DATE	1998-04-00
NOTE	58p.; For related reports, see UD 032 544-548.
AVAILABLE FROM	Consortium for Policy Research in Education, Graduate School of Education, University of Pennsylvania, 3440 Market Street, Suite 560, Philadelphia, PA 19104-3325; telephone: 215-573-0700; Children Achieving Challenge, c/o Greater Philadelphia First, 1818 Market Street, Suite 3510, Philadelphia, PA 19103-3681; telephone: 215-575-2200; fax: 215-575-2222.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Accountability; Achievement Tests; *Educational Change; Educational Objectives; Elementary Secondary Education; Organizational Development; Program Evaluation; *Standards; Tables (Data); *Teacher Attitudes; Teacher Surveys; Test Results; *Urban Education; Urban Schools
IDENTIFIERS	*Philadelphia School District PA; Reform Efforts

ABSTRACT

This report explores issues related to accountability in the context of Children Achieving, the school reform effort of Philadelphia (Pennsylvania). The accountability system begins with content standards in English/language arts, mathematics, science, and the arts. The Stanford-9 Achievement Test has been designated to assess how students are progressing under the new reforms. A Professional Responsibility Index has been developed to provide each school with a performance target that reflects expected improvements. Another aspect of the accountability system is the Keystone Schools Program, which allows the Superintendent to reconstitute any school deemed academically distressed and then reopen it under strict supervision. Performance goals have also been set for the Superintendent and his Cabinet. Teacher observation forms, and promotion and graduation requirements are other aspects of the accountability system. Data from the Children Achieving evaluation suggest that teachers have felt that they had little time to prepare or respond to the new approach, although almost all were aware of the standards by the spring of 1997, and almost all saw them as potentially beneficial to students. District support for standards-based instruction was thin and slow, and the Performance Responsibility Index was not well understood. Teachers often felt that they were being held responsible for results that are beyond their control. However, the implementation of the Stanford-9 did impact teaching practice, even though teachers did not believe it really reflected the new standards or their curricula. Although it is too early to tell how the new system will affect student achievement, recommendations are made to improve the understanding of the accountability

+++++ ED423342 Has Multi-page SFR---Level=1 +++++
system and its acceptance. (Contains 4 charts, 11 tables, and 6 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 342

96

CHILDREN ACHIEVING:
PHILADELPHIA'S
EDUCATION REFORM
PROGRESS
REPORT SERIES
1998-1999

The Accountability System: Defining Responsibility for Student Achievement

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Theresa Luhm
CPRE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

prepared by

CONSORTIUM FOR POLICY RESEARCH IN EDUCATION
RESEARCH FOR ACTION
OMG CENTER FOR COLLABORATIVE LEARNING

for

THE CHILDREN ACHIEVING CHALLENGE

BEST COPY AVAILABLE

97

ABOUT *the* CONSORTIUM *for* POLICY RESEARCH *in* EDUCATION

The Consortium for Policy Research in Education (CPRE) unites five of the nation's leading research institutions in an exciting venture to improve student learning through research on education policy, governance and finance. Members of CPRE are the University of Pennsylvania; Harvard University; Stanford University; the University of Michigan; and the University of Wisconsin-Madison.

CPRE conducts research on issues such as:

- Education Reform
- Student and Teacher Standards
- State and Local Policymaking
- Education Governance
- School Finance
- Teacher Compensation
- Student Incentives

Consortium for Policy Research in Education
Graduate School of Education
University of Pennsylvania
3440 Market Street, Suite 560
Philadelphia, PA 19104-3325
215 573 0700, ext. 0

ABOUT RESEARCH *for* ACTION

Research for Action (RFA), a non-profit educational research and evaluation group located in Philadelphia, works with public school districts, educational institutions, and community organizations to develop programs and policies that provide equitable educational opportunities and outcomes for all students. RFA conducts research about how to ensure equitable educational opportunities and outcomes for all learners. Because we value collaborative consideration of information, issues, and ideas, we also work with educational institutions and community organizations to use inquiry processes to build coalitions and to create agendas for action that are purposeful and socially just.

Research for Action
International House
3701 Chestnut Street
Philadelphia, PA 19104
Phone 215 823 2500
Fax 215 823 2510

ABOUT *the* OMG CENTER

The mission for the OMG Center for Collaborative Learning is to support organizational learning and problem-solving through consultation, applied research and professional education. Center staff draw upon recent and established theory within the paradigms of systems thinking, action research and organization behavior. The Center works with individual organizations and inter-organizational partnerships, primarily in the non-profit and philanthropic sectors. Its main focus is on urban and social policy issues; the majority of its activities are directed at finding innovative solutions to problems of community deterioration, poverty and discrimination, environmental quality, and human capital.

OMG Center for Collaborative Learning
1528 Walnut Street, Suite 805
Philadelphia, PA 19102
Phone 215 732 2200
Fax 215 732 8123

The Accountability System: Defining Responsibility for Student Achievement

THERESA LUHM

ELLEN FOLEY

TOM CORCORAN

Consortium for Policy Research in Education

April 1998

©CPRE, University of Pennsylvania

PREFACE

In December of 1995, the *Children Achieving* Challenge charged the Consortium for Policy Research in Education (CPRE) and its partners, Research for Action, OMG Center for Collaborative Learning and the Philadelphia Writing Project with the evaluation of *Children Achieving*, Philadelphia's school reform initiative. Research began in January 1996 and will continue through December 2000.

During the 1996-97 school year, the evaluation team conducted qualitative research in 21 schools, 14 clusters, interviewed District officials, and administered a District-wide survey of teachers. Drawing on this data, a series of five reports have been drafted. They include:

- Restructuring Student Supports: Redefining the Role of the School District
- Guidance for School Improvement in a Decentralizing System: How Much, What Kind and From Where?
- Making Sense of Standards: Implementation Issues and the Impact on Teaching Practice
- The Accountability System: Defining Responsibility for Student Achievement
- Technical Report on the Results of a Survey of Philadelphia Teachers

These reports are available through CPRE (215) 573-0700 extension 0 or through the *Children Achieving* Challenge (215) 575-2200.

CONTENTS

I. Introduction . . . 3

Accountability and *Children Achieving* . . . 4

The Professional Responsibility System . . . 6

II. Elements of the *Children Achieving* Accountability System . . . 7

Standards . . . 7

Stanford-9 Achievement Test . . . 9

Performance Responsibility Index . . . 12

Keystoning . . . 15

Cabinet Performance Goals . . . 16

Teacher Observation Form . . . 17

Promotion and Graduation Requirements . . . 20

Timeline . . . 21

III. System Response to the New Accountability Measures . . . 23

IV. The Results of the First Accountability Cycle . . . 31

Controversy over the Gains . . . 35

Identifying Failing Schools . . . 39

Controversy over Keystoning . . . 40

V. Preliminary Assessment of the System . . . 42

VI. Recommendations . . . 49

VII. References . . . 50

CHARTS

1. Proposed Student Promotion Policies 20

2. Timeline for the Implementation of the Standards and Accountability System 22

3. Principles of an Effective Accountability System 42

4. Effects of the Accountability System on Various Stakeholders 44

TABLES

1. Performance Levels on the SAT-9 10

2. Key Elements of the Performance Responsibility Index 12

3. Calculating a School's Growth Target for the PRI 14

4. Content Analysis of the Vision of Good Practice in the Teacher Observation Form 18

5. Teachers' Perceptions of the Standards 24

6. Teachers' Beliefs about Students' Abilities to Meet the Standards in Various Subjects 27

7. Time Spent on Review 30

8. SAT-9 Achievement Test Results, by Subject and Cluster: Elementary, K-8 and Middle Schools 32

9. SAT-9 Achievement Test Results, by Subject and Cluster: High Schools 33

10. PRI Scores, by Cluster 34

11. Performance Comparison of Two Schools 39

INTRODUCTION

Educators and policymakers share a common desire to dramatically increase the numbers of students achieving at high levels. This desire is the motivation behind most current attempts to improve the quality of public education. To accomplish this goal, many policymakers believe that schools and teachers must be held more accountable for the performance of their students. They believe that only stronger accountability will generate incentives sufficiently powerful to motivate teachers to improve classroom practice, to focus schools and districts on student outcomes, and to overcome the low standards, inertia, incompetence, fragmentation, and bureaucracy that have plagued public education and undermined previous attempts at reform.

The emphasis on accountability for performance puts the focus on student behavior and achievement (test scores, graduation rates, attendance, discipline, etc.) and represents a shift away from long-used indicators of school quality such as degrees held by the teaching staff, curriculum, special services such as libraries and guidance counseling, equipment, and facilities. Such indicators are still used for school accreditation. However, most contemporary policy makers view them as inadequate measures of quality, believing that they do not focus attention on what matters most—student performance (Ladd, 1996). While these indicators may not create the incentives needed to drive improvements in classroom practice and student performance, this should not lead to the conclusion that inputs are unimportant. To the contrary, research makes it quite clear that the presence of adequate human and instructional resources in schools is a necessary although not sufficient condition for improving performance (Hedges, Laine & Greenwald, 1994).

Contemporary thinking about accountability in education also favors systems which rely on well-defined uniform standards and objective measures, such as test scores, over systems relying on human judgments of quality, such as accreditation or peer review. In some cases these two approaches are combined, and test scores and other quantitative measures are used to identify schools that then undergo a qualitative review and analysis by professional educators. But the primary weight of accountability rests on the objective measures of student performance which are thought to be more important, more reliable, fairer, and less subject to manipulation.

Thus, state and local policy makers are adopting external objective accountability systems designed to generate incentives powerful enough to prompt educators to change their behavior in order to improve student performance. Is the current technology of assessment sufficiently reliable and valid to carry the weight of such accountability systems? Will such high stakes accountability systems stimulate more productive behavior by school staffs? Will they lead schools into processes of continuous improvement that can be sustained in the long-term? Will they stimulate desirable changes in classroom practice? These questions are matters of considerable debate, and there is no consensus at present about the answers. This paper will explore these questions in the context of Philadelphia's school reform effort, *Children Achieving*.

Accountability and Children Achieving

The principal architect of *Children Achieving* is Superintendent David Hornbeck who brought the plan to Philadelphia and was hired to implement it. Because the incremental reforms of the past have proved inadequate, Hornbeck believes that the stakes must be raised, and that radical, comprehensive changes are required to help urban children achieve at high levels. His vision was adopted by the Philadelphia Board of Education as the ten components of *Children Achieving*. In his own words, Hornbeck explains:

There are ten components of comprehensive, systemic change that must occur over the next five years if we are to have the learning environment in schools and communities in which large majorities of all children demonstrate high achievement. In broad terms, they are:

1. *We must behave as if we believe that all students will learn at high levels...*
2. *Standards-based reform will drive the system...We must set standards, have new assessment strategies, and develop new incentive systems for both adults and students in the system.*
3. *Decisions will be made at the school level.*
4. *Staff development is critical to improved performance.*
5. *Early childhood support is less expensive and more effective*
6. *Community services and supports can make the difference between success and failure.*
7. *Adequate technology, instructional materials, and facilities are necessary to learning...*
8. *Strong public engagement is required.*
9. *We must have adequate resources and use them effectively...*
10. *We must do all of these nine components.*

However, the "theory of action" underlying this vision for *Children Achieving*—that is, the dynamic relationship among the elements that will lead to improved performance—is not explicit in this list of components. Based on other statements made by the Superintendent, the theory of action underlying *Children Achieving* would seem to be:

Provided high academic standards and strong incentives to focus efforts and resources, greater control over school resource allocations, organization, policies, and programs, adequate funding and resources, more hands-on leadership and high-quality support, better coordination of resources and programs, restructured schools in order to support good teaching and encourage improvement of practice, rich professional development of a person's own choosing, and increased public understanding and support, the teachers and administrators of the Philadelphia schools will develop, adopt, or adapt instructional technologies and patterns of behavior that will help all children reach the District's high standards.

The critical "drivers" in this theory are the standards and the incentives embedded in the newly adopted accountability system. In October 1996 the Philadelphia Board of Education approved a plan to hold administrators and teachers professionally responsible for the achievement of students. The *Children Achieving* Action Design includes the rationale for implementing the system of accountability:

Trying hard is not good enough, either for those who work in the system or, ultimately, for students. Under the High Expectations component, the District has outlined what the systemwide standards for all Philadelphia's public school students must be if we are to achieve our vision. There are, however, two

other primary features of a high expectations, performance-driven system that the district must implement: assessment and accountability. Weaving the web between standards, assessment, accountability and support of good teaching and learning is the central feature of systemic change.

Our high expectations, performance-driven system must use incentives that apply to both students and staff. Intrinsic incentives will always be powerful. For a professional educator, the satisfaction that comes from succeeding with a youngster with a history of failure or helping a talented student stretch her horizons is unparalleled. Nonetheless, extrinsic incentives such as financial and professional rewards for both school-based and non-school based personnel also have a role in a meaningful accountability system.

As the *Children Achieving* Action Design articulates, the focus on results and their publication and the use of rewards and sanctions—extrinsic incentives—are central to Philadelphia's accountability system. They are the engine that will drive school improvement across the District. According to the theory, this will occur in one of two ways: either teachers will teach better because they desire the reward (either cash or public recognition) obtained for higher test scores, or teachers will teach better because if student performance fails to improve, they will be subject to various sanctions.

From the perspective of the Superintendent and the Board of Education, intrinsic rewards are simply not enough; extrinsic rewards are needed to change people's behavior and ultimately the performance of the system. However, it is important to note that *Children Achieving* offers no definition of what "teaching better" entails. It does not prescribe a particular approach to instruction. Rather, by decentralizing decision making, the reform plan leaves it up to schools, small learning communities and teachers to decide how to improve student achievement. The accountability system emphasizes standards and outcomes, but does not spell out the steps necessary to obtain them.[1] The premise is that given support and the freedom to make decisions, educators will be motivated by the rewards and/or sanctions to aggressively seek and use better methods and programs of instruction, or if they do not exist, to invent them.

The underlying assumption of the accountability system in Philadelphia is that if educators work harder and smarter and have adequate resources and supports, student achievement will improve. Under this system, teachers and administrators must assume a larger burden of responsibility for the performance of their students. School staff are being asked to help all students, not just the academically motivated, reach high standards. They are expected to help students acquire deep understanding of content in the core subject areas and to integrate and apply that knowledge to real problems. Educators at all levels of the system are being asked to master new skills, take on new responsibilities, and change their practice in order to meet the needs of a diverse and disadvantaged student population. Is this reasonable? Hornbeck argues that it is:

Professional responsibility is not about punishing teachers. It is not about their being responsible for the social difficulties that so many children bring to school with them, for the shortage of resources that we have in our schools, or for any factor that is beyond their control. It is, however, a recognition that teacher and administrator performance plays an important role in the success or failure of our students.

[1] For a more detailed discussion about how *Children Achieving* is impacting instruction see Simon, E., *Making Sense of Standards: Changing Instructional Practice in the Context of the Children Achieving Reform.*

The Professional Responsibility System

This paper describes the accountability system being implemented in the Philadelphia schools, the initial responses of educators to the system and some of its components, its initial impact on teaching practice, and its connection to the District standards. The key questions which will be addressed include:

- What is the system of accountability?
- How well has it been designed and implemented?
- How have various stakeholders responded to it?
- How well does the accountability system meet contemporary standards of quality?

Of course, the ultimate test of the accountability system is whether it leads to sustained improvements in student achievement. However, it is simply too early to make such an assessment of the new system in Philadelphia since 1996-97 was the first year of the first two year accountability cycle. In 1996-97 sixteen of the 22 school clusters were in their first year of implementing the components of *Children Achieving*. The standards, the rewards and the sanctions associated with the Professional Responsibility System were new to the teaching staff, and the incentives associated with the system had not had time to affect policies and practices in the schools. Indeed, it is probably the case that the idea of rewards was simply an abstraction to many teachers during the 1996-97 school year, and it is likely to remain so until they have been distributed for the first time. The full motivating force of rewards may not come into play until they have been distributed and are in the budget. The power of the sanctions was more apparent as a result of the District's effort to reconstitute two high schools; however, systemwide rewards and sanctions will not be distributed until the summer of 1998. Therefore judgements about the impact of the new accountability system on staff behavior and student performance will have to be addressed in the future when further data have been gathered and analyzed.

ELEMENTS *of the* CHILDREN ACHIEVING ACCOUNTABILITY SYSTEM

The accountability system in Philadelphia is comprehensive and complex, affecting people at all levels of the District from the Superintendent to students in the classroom. Some elements have been put into place and others are still being designed. This section of the paper will outline these elements and discuss who they affect and how they are affected.

Standards

Accountability in Philadelphia begins with standards. Starting in late 1995, the School District of Philadelphia, assisted by the Philadelphia Education Fund, convened groups of teachers, parents and others to develop content standards in English/language arts, mathematics, science and the arts. Drawing heavily on content standards created previously by professional groups such as the National Council of Teachers of Mathematics (NCTM) and the National Council of Teachers of English (NCTE), Philadelphia's Standards Writing Teams drafted content standards in the aforementioned areas, integrating several "cross-cutting competencies"—citizenship, technology, multicultural competence and problem solving, for example—in all the disciplines.

The initial set of draft standards in English/language arts, mathematics, science and the arts was reviewed by Standards Review Teams made up primarily of teachers and other educators from the District. A second draft, incorporating revisions suggested by the Review Teams, was then distributed to all teachers in August 1996. After public meetings were held on the standards in all 22 clusters, the Content Standards, Benchmarks and Performance Examples were adopted by the Philadelphia Board of Education in December 1996. A similar process was followed to develop the Content Standards, Benchmarks and Performance Examples in Health and Physical Education, Social Studies, and World Languages. Standards in those disciplines were adopted by the Board of Education in July 1997.

The authors of Philadelphia's Content Standards were careful to note that the standards were not intended to "dictate how material should be taught and what curriculum should be used." However, leaders of the Philadelphia standards movement also knew that they had to provide some guidance to teachers and other school personnel to help them understand what standards-based instruction looked like and what they had to do to implement the standards. In July 1996 the District offered a three-day professional development session on standards-based instruction that was attended by many teachers. Additionally, the District issued Standards Curriculum Resource Guides in English/Language Arts, Mathematics and Science in September 1996 for each of three grade ranges, K-4, 5-8 and 9-12. The Curriculum Resource Guides were intended to help teachers make a transition from the Standardized Curriculum, Instructional Planning Guides, and Marking Guidelines that had been issued by the previous District administration to the use of the new content standards that were then under development. In addition, the District developed a Resource Guide for Standards-Based Assessment and Instruction in February of 1997 which linked the content standards with the District-wide assessment, the Stanford-9 Achievement Test, and offered a few examples of lessons incorporating the standards.

While the standards documents and Standards Curriculum Resource Guides provided teachers with examples of standards-driven teaching and assessment strategies during the roll-out in the 1996-97 school year, teachers demanded more specific materials to help them understand more fully what they should be doing in their classrooms. These concerns led to lengthy discussions among the central office

staff about their role in a decentralized environment and what the character of any guidance documents or curriculum materials developed in support of the standards should be. While most of the central- and cluster-level educators involved in these debates felt that some supports were necessary, they disagreed about who should develop them and what they might entail. Some of the options discussed were:

- central office-initiated development of curriculum frameworks which would detail grade-by-grade objectives and sequencing of the content standards;
- central office-initiated development of a document that would act as a bridge between the old, standardized curriculum which was initiated by the previous Superintendent, and the new system of the content standards;
- time and resource support for school-initiated development of curricula based on the content standards;
- central office-initiated development of performance standards (how to measure what students know and are able to do) aligned with the content standards,
- central office-initiated development of model units of study;
- time and resource support for school-initiated development of units of study; and
- central office provision of access to schools to units of study developed by schools/educators outside the School District of Philadelphia.

In January 1997 the decision was made to develop detailed curriculum frameworks in four major subject areas (English/language arts, mathematics, science, and social studies) for use District-wide. This decision had been delayed by debates over which level of the system should assume responsibility for developing resources in support of standards-based instruction. While teachers were clamoring for more support, central office staff were debating whether the provision of more specific curricular guidance would rob schools of the opportunity to develop their own curricula and contradict a central tenet of *Children Achieving*, that decisions should be made at the school level as much as possible.

Even after deciding to develop frameworks, discussions continued about their appropriateness and their content in a standards-based environment. Should these documents delineate what students should know at the end of each grade? Some central office staff felt that they should; others thought this would contradict the developmental approach to instruction underlying the standards. From this perspective, since children develop academically at different rates, they should be allowed different amounts of time to reach the same standards.

Another concern about the curriculum frameworks centered around units of study. Some District officials believed that standards-driven instruction should be organized around interdisciplinary themes that encompass a number of standards and end with a culminating task. However, the Curriculum Frameworks that were developed do not directly promote the development of such units of study. The Frameworks, extensive detailed documents with grade-by-grade sequencing of the content standards, were published in January 1998. As of this writing, neither performance standards nor opportunity-to-learn standards have been adopted. However, District officials view the newly developed school support process and the proposed draft graduation and promotion policies as one level of performance and opportunity-to-learn standards. Only time will tell whether the combination of the standards documents and the frameworks will provide sufficient guidance for school staffs or small learning communities to develop, select, or adapt curricula, and whether they are sufficiently aligned with the assessment to have the powerful cumulative effect on teaching and learning that District leaders envision.

Stanford-9 Achievement Test

During the 1994-95 school year, School District officials introduced the Stanford-9 Achievement Test (SAT-9) to assess how students were progressing under the new reforms. The SAT-9 is a criterion-referenced assessment composed of selected and constructed response items. The test is linked to voluntary national standards developed by professional organizations such as the NCTM and the NCTE. Before deciding on the SAT-9, District officials seriously considered using the New Standards Project's Reference Examinations but decided against them because of their cost, the lack of examinations in all of the core academic subjects, and the lack of data about their technical adequacy. One of the reasons the SAT-9 was chosen over others is that the publisher, Harcourt Brace, agreed to align the SAT-9 with the District's own standards as they were adopted. In an April 7, 1997 news release, Rich Maraschiello, a research associate in the District's Office of Assessment, explained that the linkage to standards makes the SAT-9 more challenging. He said, "The standards require that children do a lot more critical thinking and problem solving. In addition to multiple choice questions, the SAT-9 contains many open-ended questions that demonstrate the student's proficiency with these skills as they apply in the basic subjects—reading, math, and science."

An additional reason District officials gave for choosing the SAT-9 is that it is a criterion-referenced test. With most achievement tests, scores are compared to national norms but the SAT-9 also compares student performance to established levels of achievement that represent levels of competency in a subject. Student scores are categorized by performance levels—Below Basic, Basic, Proficient, or Advanced in each subject area. Scores on the SAT-9 show how close a student is to achieving a proficient level as well as how he or she compares to other students across the nation. Because a large percentage of Philadelphia students score in the Below Basic category, District officials asked the test publisher to break down that level into three separate categories so that schools would be credited with the progress made by students who were improving but not yet performing at the basic level. The resulting six categories of performance being used in Philadelphia are summarized in Table I below[2].

[2] Table taken from School District of Philadelphia document "Definitions and Policy for SAT-9 Administration."

TABLE 1

Performance Levels on the SAT-9

SAT-9 Performance Level	Definition	Percent of Students Achieving Level Nationally
Advanced	Superior performance beyond grade-level mastery. High school students achieving at this level show readiness for advanced academic courses, advanced technical training, or career-oriented employment.	Nationally, fewer than 10% of students achieve the advanced level on the SAT-9.
Proficient	Solid performance, meaning students are ready for the next grade. At high school, this level reflects competency in the body of subject-matter knowledge and skills that prepares students for responsible adulthood, productive work and further education.	Nationally, fewer than 25% of students achieve the proficient level at most grades and subjects. This figure is lowest in 11th grade math and science, where fewer than 10% of students achieve at this level.
Basic	Partial mastery of the knowledge and skills that are fundamental for satisfactory work. At the high school level this is higher than minimum competency skills.	Nationally, more than one-third of students achieve the basic level on the SAT-9, except for 11th grade math (19%) and science (26%).
Below Basic III	Inadequate Mastery	Nationally, the proportion of students scoring at the below basic level on the SAT-9 ranges from 25% in 4th grade mathematics and science to about 70% in 11th grade math and science.
Below Basic II	Little mastery	
Below Basic I	Very little mastery	

During the 1996-97 school year, the SAT-9 was administered in grades 2, 3, 4, 6, 8, 9 and 11 in reading, math and science. The test was first administered to the students in grades 4 and 8 in the 67 schools in the original six clusters during the spring of the 1994-95 school year. The test was administered District-wide in the two subsequent years. District officials require schools to administer the test to all students. The only students exempted are those who are classified as severely and profoundly impaired, as trainable mentally retarded or those in ESOL at Level 1. Any student who does not attempt both the open-ended and the multiple choice section of a content test is given a score of zero which affects how a school performs on the District's accountability index (discussed on the following page). This is intended to promote the participation and achievement of all students and to ensure that school administrators do not "inflate" school performance by testing only those students who they believe will perform well.

Although the SAT-9 measures individual student achievement, its primary purpose in Philadelphia is to hold schools, teachers, principals, the Superintendent and his Cabinet accountable. The scores on the assessment are used as one part of a numerical index that is used to gauge the progress schools are making toward improving student achievement. Ultimately, District officials would like to use additional measures of student achievement to judge progress. The goal is to expand and strengthen the subjects included and the array of tools that principals, parents and teachers use to gauge how well schools and students are performing. In the future, this could include the use of student portfolios and exhibitions as well as the development of a school quality review process. While District officials are working on developing these new assessments, none have been adopted systemwide.

The adequacy of the SAT-9's alignment with the Philadelphia standards was an issue of some controversy during the 1996-97 school year because some teachers did not believe that the test matched their curriculum. As a consequence, District officials had the assessment and the standards reviewed by District content specialists. The results were mixed. According to one District official, the specialists found good alignment in English/language arts, while in mathematics, he noted that examining "the SAT-9 actually informed us about the gaps in the standards." He also acknowledged that the mathematics test requires a high level of literacy which is somewhat problematic in Philadelphia because reading proficiency is very low. He also described the alignment between the standards in science and the assessment as "not good." The problem was not that the standards and the SAT-9 didn't cover the same material, but that the material was not covered at the same grade levels. He questioned whether the SAT-9 was "developmentally appropriate" in science.

Based on the results of this review, District officials have worked with the test publisher to create some new items for the test to be administered in the spring of 1998. Some of these items were included on a pilot basis on the assessment administered in 1997. Whether or not these new items result in better alignment with the standards remains to be seen. They have also developed and piloted other new items to cover multicultural contexts which officials believe will be "ground breaking," since such items are not currently in use anywhere. This would expand the SAT-9 to cover one of the cross-cutting competencies included in the Philadelphia content standards.

Performance Responsibility Index

On October 21, 1996 the School Board adopted the framework for the Professional Responsibility Index (PRI). The goal of the PRI is to provide each school with a performance target that reflects expected improvements in the following areas:

- math, reading, and science scores of students in grades 4, 8, and 11 on the SAT-9;
- promotion to the next grade level in elementary and middle schools;
- the proportion of 9th grade students graduating from high school in four years; and
- student and staff attendance.

Using 1996 as the baseline, the District set long- and short-term targets for each school on an index comprised of these indicators. All schools have the same long-term goal; within 12 years—by 2008—each school, and the District as a whole, should achieve a score of 95 out of 120 on the performance index. In the interim, each school must move one-sixth of the way closer to the 2008 goal in each two-year period. At the end of each of these two-year intervals, the baseline and the targets are reset for each school. To check progress, the District issues an annual Report Card for the system as a whole and for each school[3].

The SAT-9 scores count for 60 percent of the overall index score. Reading, math and science count for 20 percent each. The student promotion/persistence rate[4] counts for another 20 percent. Student and staff attendance scores count 10 percent each. In addition to meeting its performance target, each school must achieve at least a ten point drop in the percentage of students achieving below the basic level in reading, math and science on the SAT-9.

TABLE 2

Key Elements of the Performance Responsibility Index

- affects all schools
- targets based on common long-term performance goal for all schools
- two-year performance cycles
- targets set based on linear progress over two-year performance cycles
- based on statistical indicators
- includes non-cognitive indicators
- includes rewards and sanctions and assistance for distressed schools

[3] How each variable in the accountability index is calculated is discussed later in this report in the section Controversy over the Gains.

[4] The promotion rate is the percentage of students in grades one through eight that are promoted either by policy or by exception. Students who are assigned to the next grade are not counted as promoted. The persistence rate measures the proportions of first-time 9th graders who graduate from Philadelphia School District high school four years later.

The rationale for using an index instead of relying solely on the SAT-9 scores is that the District wants schools to raise the performance of all students, not just those who currently show up on the days the test is administered. To achieve this policy goal, they have built into the index other indicators—attendance, promotion, graduation—to promote school attention to them. This means that schools cannot raise their performance on the SAT-9 simply by pushing low performers out. This is reinforced by the inclusion of students in the assessment system who were previously untested, such as some categories of special education and ESOL students.

Starting in 1998, schools will receive rewards, assistance or sanctions based on whether they did or did not reach their two-year performance targets. These include:

- Schools that exceed their targets will receive public congratulations and will receive an award of $1,500 for each teacher and $500 for each of the other staff members.
- Schools that meet but do not exceed their target will be publicly recognized for their accomplishment, but will not receive awards.
- Schools that improve beyond their baseline (1996) scores but fall short of their targets will receive help from a team of educators who will review school information, assess school resources and help find ways for schools to improve.
- Schools that drop below their baseline scores will receive help from support teams. Further, administrators and teachers in those schools will face close scrutiny through the District's rating system. Those who receive poor evaluations may be denied wage and step increases or, if problems continue, may be terminated.
- Schools that fail to meet their short-term goals for two accountability cycles (four years) in a row also will receive help and close evaluation. They may also face reconstitution, a process that could result in the transfer of up to 75 percent of staff.

Assumptions Underlying the PRI

What assumptions underlie the PRI? School District officials decided on a 12-year time period for all schools to achieve at least a score of 95 on the index. This is based on the notion that by 2008 all students will have experienced the effects of the *Children Achieving* reforms from kindergarten through 12th grade. This reasoning led to the adoption of the following method for determining each school's threshold, the score each school must meet or exceed every two years to be rewarded and avoid sanctions:

Since there are 6 two-year cycles in the 12-year period, to make steady progress toward the long-term goal, schools must gain one-sixth of the difference between the first baseline score and the target of 95. For example, if a school's baseline index in 1996 was 69.2, that score subtracted from 95 (25.8) represents the growth the school needs to achieve over 12 years. To find the school's two-year growth target, the total growth needed (25.8) is divided by 6. Therefore the school's growth target is 4.3. To calculate the school's performance target for 1998, the growth target (4.3) is added to the school's 1996 baseline index score, so the 1998 performance target becomes a score of 73.5 on the PRI.

This calculation is illustrated in Table 3 on the next page.

13

TABLE 3

Calculating a School's Growth Target for the PRI

Performance Target

1

Baseline Index	
Reading	56.4
Mathematics	55.4
Science	46.4
Promotion	90.1
Enabling	97.7
Baseline	69.2

Total Growth Needed Over 12 Years (96-Baseline)

2 25.8

Growth Target for 2 Years (Total Growth/6)

3 4.3

Performance Target (Baseline + Growth target) and a 10 point reduction in the percent of students below the basic level

4 73.5

In addition to increasing the total score on the PRI, school staff must also reduce the number of students scoring below the basic level in reading, math and science by ten percentage points in order to meet their performance targets. This ensures that schools must succeed with their lowest performing students as well as those at the top. If this was not part of the index, school staff could concentrate their efforts solely on those students most likely to move from one level to the next—for example, from proficient to advanced.

District officials argue that the Performance Responsibility Index is designed to put all schools on a "level playing field." The point is not to compare schools to each other, but to put all schools on a path "toward a common standard of excellence." They are being compared against themselves rather than each other.

Keystoning

As mentioned in the previous section, the harshest sanction which District officials can level against a school under the PRI is reconstitution. Schools which fail to meet their short-term goals for two rating periods (four years) in a row may have up to 75 percent of their staff force transferred. During contract negotiations between the School District of Philadelphia and the Philadelphia Federation of Teachers in August 1994, both parties recognized that significant action should be taken to address the problems of academically distressed schools. They agreed on a policy known as the Keystone Schools Program which allows the Superintendent to reconstitute any school deemed academically distressed and reopen it as a Keystone school subject to the following conditions:

- Notification of the intent to reconstitute will be given no later than February 15 of the preceding school year;
- The principal of the school will be designated no later than March 15 of the preceding school year;
- A Comprehensive Education Renewal Plan for the Keystone school will be developed by a committee consisting of the designated principal, eight senior career teachers appointed by the Federation, and at least one parent representative from each grade;
- A joint committee comprised of four representatives of the School District and four representatives of the Federation will monitor and support the implementation of the Keystone Schools Program;
- Any staff member assigned to the reconstituted school who wants to be reassigned or is reassigned to a different location will be allowed to do so and will be treated as a forced transfer;
- Up to 25 percent of the new staff may be selected by the principal from among existing staff pursuant to the criteria developed by the principal in conjunction with the Federation;
- The remaining 75 percent of the new staff and staff in succeeding years will be filled from other than teachers previously assigned to the school;
- Each Keystone School is required to submit a report at the end of each school year to the Joint Committee documenting improvements in student achievements.

Although the PRI had not yet been developed when the PFT and the School District agreed on the Keystone Schools program, reconstitution became available as a sanction for the Superintendent during February 1995.

Cabinet Performance Goals

The Superintendent and his Cabinet members are also being held accountable for improving student achievement levels. They must meet specific performance goals (negotiated with the School Board) each year or the Board can penalize them by withholding up to five percent of their salaries. In the agreement for the 1996-97 school year, a total of 50 percent of their performance was measured according to student achievement results on the SAT-9. The other 50 percent was judged according to a set of "enabling goals." These include goals in the following areas:

- standards, curriculum and instruction;
- accountability and assessment;
- student support services;
- facilities and technology;
- school safety;
- local decision making;
- public engagement; and
- resources.

According to the January 13, 1997 letter of agreement between the Superintendent and the School Board, any bonus or penalty for the Superintendent is based directly on both student achievement and the enabling goals. With respect to the members of the Cabinet, the Board sets aside an amount of money which the Superintendent then allocates among the Cabinet members based on his evaluation of each member's performance.

One-half of any bonus for each central office Cabinet member depends on whether systemwide student achievement exceeds the systemwide performance target. The balance depends on the quality with which each central office Cabinet member met the goals agreed upon individually between the Cabinet member and the Superintendent. For each cluster leader, one-third of the bonus depends on systemwide student achievement; one-third on student achievement within the cluster leader's own cluster; and one-third on goals agreed upon between the cluster leader and the Superintendent. With respect to penalties, the Cabinet faces potential income penalties in the same proportions and the same basis as rewards, but only up to five percent[5].

[5] For the 1997-98 school year, the School Board has proposed that any bonus or penalty for the Superintendent and Cabinet be based solely on student achievement.

Teacher Observation Form

During the 1996-97 school year, a new evaluation form was adopted by the School District of Philadelphia for use by principals when conducting classroom observations. The intent of the new form is to hold teachers accountable for implementing the new content standards in their classrooms. According to District officials, the new form represents the first time student performance has been included as part of the teacher rating system. The top of the form contains information about the observation: name of teacher, school, date and time of observation, room number and subject being taught, etc. There is a place for the principal to identify the teacher's small learning community and a place to indicate the students enrolled in the class and the number who are actually present, implying that the attendance percentage is a matter of concern. The principal is also required to check a box to indicate whether the lesson involved the whole class led by the teacher, small groups working under the teacher's direction, small groups working independently, or students working as pairs or individuals, acknowledging that a lesson's overall characteristics might be influenced by the way students are grouped.

TABLE 4

Content Analysis of the Vision of Good Practice in the Teacher Observation Form

Messages Being Communicated by Items in the Form	No. of Items
1. Engage in developmentally appropriate activities that help students construct knowledge and meaning = *incorporate ideas from professional development; draw on prior knowledge and interests of students; accommodate different learning styles and needs; use available technology, equipment and materials; use a variety of assessment strategies.*	12
2. Establish a positive instructional climate = *establish attractiveness and functionality of the classroom environment; focus on student-teacher and student-student interactions; establish behavior guidelines, respect for others, and high expectations.*	10
3. Hold high expectations for all = *promote achievement and participation of all students, communicate high expectations, hold students accountable for behavior.*	6
4. Focus on the standards = *connect lessons to the District-adopted standards, or otherwise to the goals of the SLC, school or District; relate teaching approach to recent professional development.*	6
5. Provide authentic instruction = *engage students in higher-order thinking and substantive conversation with each other and the teacher; have students reflect on their work, and make connections to other areas of the curriculum or to real-world experiences.*	6
6. Reach beyond the classroom = *develop positive interactions with others in the school and community, including parents; relate content to real-world experience; relate basic skills to broader concepts; include critical thinking and problem solving.*	5
7. Assess students = *use a variety of assessment strategies, give a variety of feedback; demonstrate progress towards instructional goals*	3

Single mentions included: the teacher's knowledge of the content, using an instructional plan linked to the plan of the SLC, and using technology.

This background information is followed by a list of 33 items on which the principal must rate the teacher's performance on 4-point scale, ranging from "no evidence" to "positive and sustained presence." On the form, these 33 items are grouped into four categories: personality, preparation, technique and student reaction. The items in the personality section concern the teacher's overall attitude—relating to the broader community, dealing with diversity issues, using common sense. The preparation section includes statements about the teacher's knowledge of content, standards, and pedagogy; about planning effectively; about keeping good records; and about being aware of school and community resources. The items in the sections on technique and student reaction seem to overlap considerably, and understandably, since the processes of learning and teaching are so intertwined.

But, upon closer examination, there are less obvious messages about instruction which are being communicated through the items on the observation form. These messages go beyond the four categories on the form. The messages and the number of items that allude to them are enumerated in Table 4. This analysis reveals a point of view about instruction that is not only consistent with the standards but reflects a student-centered, constructivist approach. While this point of view is generally consistent with national standards of good practice, the inclusion of such criteria on the teacher evaluation form seems somewhat inconsistent with the stated district policy that schools and small learning communities are free to select their methods to prepare students to master the standards. For example, the observation form would appear to rule out heavy reliance on direct instruction.

Promotion and Graduation Requirements

Superintendent Hornbeck has stated repeatedly, "Accountability must not apply to educators alone. Parents, students and the wider citizenry also are responsible." In October 1997 the School District released a draft plan designed to increase student accountability. At every level of their education, students face increased requirements for promotion and graduation. The new requirements are scheduled to take effect in the year 2000 and are to be increased yet again in 2002. The proposal is still in draft form at this time, and District officials are in the process of obtaining citizen feedback. Chart 1 gives an example of the changes that have been proposed.

CHART 1

Proposed Student Promotion Policies

Promotion from Grade 4
Current Requirements • Must pass 3 of 4 major subjects • Students can not be retained in grade 1 except by exception and no child may be retained more than once
Proposed Requirements, Effective September 2000 • Must pass language arts, math and science • Must complete a project that involves more than one subject and requires strong writing skills • Must obtain a score of at least below basic III on the SAT-9 reading and math tests or demonstrate at least third grade proficiency on District-wide reading and math assessments
Proposed Requirements, Effective September 2002 • Must pass language arts, math and science • Must complete a project that involves more than one subject and requires strong writing skills • Must complete a project demonstrating citizenship through community service • Must score at least basic on the SAT-9 reading and math tests or demonstrate at least fourth grade proficiency on District-wide reading and math assessments.

However, the new tougher promotion and graduation requirements will not be put into effect unless new supports are put in place. These new supports involve parents, teachers, students, administrators, and the greater citizenry:

- **Parents** must ensure good attendance, provide for health needs, encourage reading, support homework, teach responsibility, know school rules, and communicate with the school.
- **Professionals** must have good attendance, provide effective instruction, increase student learning, and have good classroom management. (This new standard has already been incorporated into the Performance Responsibility Index.)
- **Students** must have good attendance, make a real effort, exhibit good behavior, and show respect for others.
- The **School System** must have a support process in place in every school, and provide summer school/extended time for failing students.
- The wider **citizenry** must maintain funding for current programs, volunteer more, provide funding for summer school and intensified supports, increase workplace learning opportunities, and provide more books and computers.

While monitoring funding levels and attendance is a relatively simple task, it is not clear how District officials intend to decide whether parents are providing the support that has been requested. There has been some discussion of having parents sign a "contract" that outlines their particular responsibilities, but how this will be upheld is a difficult issue that has not yet been resolved.

Timeline

As mentioned at the beginning of this section, different elements of the accountability system have been put into effect at different times, and some components are still not in place. The timeline on the following page shows when the pieces were formally adopted in chronological order.

CHART 2

Timeline for the Implementation of the Standards and Accountability System

April-May 1995	SAT-9 administered in first 6 clusters in grades 4 and 8
August 1995	First 6 clusters formally established
December 1995	Standards Writing Teams convened; writing of standards begins
April-May 1996	Copies of draft standards distributed (English/Language Arts, Mathematics, Science and the Arts) for review; SAT-9 administered District-wide in grades 2, 4, 6, 8 and 11
Summer 1996	First 4 sets of standards reviewed by Standards Review Teams
July 1996	Four-day professional development session conducted for teams of teachers on standards-based instruction
August 1996	Second draft of above standards distributed to all teachers for review
September 1996	Standards Curriculum Resource Guides for grades K-4, 5-8 and 9-12 distributed to teachers; 16 new clusters brought on line
Oct.-Nov. 1996	PRI adopted by Board of Education; Public hearings on recommended standards held in all 22 clusters
December 1996	First 4 sets of Recommended Content Standards; Benchmarks and Performance Examples with minor revisions adopted by board
January 1997	Review copies of draft standards distributed (Health and Physical Education, Social Studies and World Languages)
February 1997	Resource Guide for Standards-based Assessment and Instruction distributed to schools; Announcement of plans announced to reconstitute Olney and Audenried High Schools
April-May 1997	SAT-9 administered District-wide for grades 2, 3, 4, 8, 9 and 11; Second draft of above standards distributed to all teachers
July 1997	Final three sets of Recommended Content Standards, Benchmarks and Performance Examples with minor revisions adopted by board; Reconstitution decision for Olney and Audenried reversed by arbitrator; Week-long, content-based professional development session conducted for teams of teachers (totaling 1,100) on content standards in English Language Arts, Mathematics and Science
September 1997	PRI scores made public (232 schools improved on the index, 77 met their targets, and 15 labeled "low progress"); A second week-long professional development session on content standards attended by 600 teachers participating in school teams
October 1997	Development of Curriculum Frameworks begins
January 1998	Curriculum Frameworks for English/Language Arts, Mathematics, Science and Social Studies distributed to all schools; SAT-9 scores adjusted to correct error by test publisher; Two schools removed from "low progress" list
April-May 1998	SAT-9 administered District-wide in grades 2, 3, 4, 7, 8, 10 and 11
August 1998	First cycle of rewards/sanctions based on PRI to be released

SYSTEM RESPONSE *to the* NEW ACCOUNTABILITY MEASURES

During the 1996-97 school year, the *Children Achieving* evaluation team conducted qualitative research in 21 schools and 14 clusters, interviewed district officials, and administered a District-wide survey of teachers. Drawing on this data, this section of the paper will discuss how various stakeholders have reacted to the implementation of the different components of the new accountability system. The first section is organized around a set of "findings" from the data.

Finding: As the components of the accountability system were introduced very quickly, teachers felt they had little time to prepare or respond thoughtfully.

District officials felt it was essential to get a baseline reading of system performance on the SAT-9 during the 1995-96 school year. They chose not to wait two years for the implementation of standards and other supports. Whatever the merits of this decision, from the viewpoint of many teachers and some other observers, elements of the accountability system have been implemented too quickly and in an inappropriate sequence. For example, the Stanford-9 Achievement Test was administered for the first time to all students in grades 2, 4, 6, 8 and 11 in the District in April and May of the 1995-96 school year. Teachers were provided with little preparation for the test as the results were to serve as the baseline data for the accountability system. It was not until December of the next school year that the core content standards were formally adopted by the School Board. The standards are supposed to be the framework teachers use to develop curriculum and improve student performance. However, the second administration of the SAT-9 took place only 4 months after the standards were adopted. On the Survey of Philadelphia Teachers, a little more than half of the teachers reported that they had had adequate time to implement standards in their classrooms. In addition, only a third reported that they had the curriculum materials needed to help students meet the content standards. The implementation simply happened too quickly for some teachers to make the necessary adjustments. As one teacher explained:

I have no problem with standards. The problem is the way they were given to us. The average teacher would like to come in September and hear, "Here are the standards, now use your creativity to implement them." But what we have is confusion. We don't even know what draft of the standards to use. And then there is the problem of different math series. We are using one and when we looked at it against the standards, we found it lacking in performance assessment. So, what are we supposed to do? We can't change everything we do in two weeks.

Overwhelmingly, teachers told researchers that they were resentful of the way the accountability system was rolled out because there was so little time allowed for preparation. One teacher summed up the feelings when she said, "They [District staff] seem to be saying that professional development is forthcoming, but in the meantime, we'll evaluate you according to what you are going to learn. How is that fair?"

Finding: By the spring of 1997, almost all Philadelphia teachers were aware of standards and saw them as potentially beneficial to students.

By the end of the 1996-97 school year, most teachers were aware of the content standards and had looked them over even if they had not read the Recommended Content Standards, Benchmarks and Performance Examples booklet. This finding is supported by both the qualitative and survey data. As shown in the table below, a large majority of teachers (83.2%) who responded to the survey believed that they understood the purpose of the content standards and a similar number of teachers (80.8%) felt the standards had the potential to benefit students.

BEST COPY AVAILABLE

TABLE 5

Teachers' Perceptions of the Standards

Survey of Philadelphia Teachers, 1997

Statement about Content Standards	Percentage of teachers who agreed with each statement
I understand the purpose.	83.2
I believe it has the potential to benefit by students.	80.8
I believe that it already has had positive effects in my school.	57.4
I believe that it already has had negative effects in my school.	22.3
I believe that it has had no effect in my school.	35.7

Clearly, not only did the vast majority of teachers believe they understood the purpose of the standards, but a majority indicated they had high hopes for the potential of standards to benefit their students. A majority felt the standards had already affected their schools positively and very few saw negative effects. About one-third did not see the standards having any effects.

What happened during the 1996-97 school year to increase the level of awareness and teachers' beliefs in the purpose and potential impact of standards? For one thing, all teachers received a copy of the Recommended Content Standards and Benchmarks for RELA, Math, and Science booklet at the end of the 1995-96 school year and had it to review over the summer. Secondly, the District offered training in the content standards to teams of teachers from each school during the summer of 1996 so that at least a core of teachers at each school had had an opportunity to learn more about the standards from knowledgeable District personnel, and during the school year, most teachers participated in some professional development on the content standards in their clusters or schools. Finally, publicity about the testing and accountability index as well as a District-wide Title I assessment effort raised school staff members' awareness of performance assessment and the accountability system.

Finding: District support to help teachers create standards-driven classrooms was thin and slow.

While most teachers responded on the survey that they understood the standards and their purpose, evidence from interviews and observations indicates that many were not sure what it meant to deliver standards-driven instruction. Unfortunately, guidance from the central office on this matter was rather weak. The initial support available to teachers from the central administration came in the form of the standards booklets describing the content standards for English/language arts, math, science, and the arts, which was issued in August 1996. Two more documents were developed and distributed later, although these did not seem as widely known to teachers in the spring of 1997. These were Curriculum Resource Guides for each school level that were structured like the standards booklet but offered additional performance examples and content ideas, and a Resource Guide for Standards-Based Assessment and Instruction which tied standards with the SAT-9 and offered a few examples of lessons incorporating the standards. The Office of Best Practices was without leadership during the 1996-97 school year, so it

did not contribute to guidance and support of practice. The new evaluation form for teachers offered another potential source of guidance, but because observations are limited, the form seemed to have had little influence on instruction in the 1996-97 school year. Beyond these materials, it was left to the clusters, the instructional improvement arm of the District, to assist school staff in understanding and implementing standards-driven instruction. There was a great deal of variation across the clusters in the professional development and assistance they provided for schools. (For further detail, see the report on decentralization.)

In sum, although teachers knew about the standards and thought they could be beneficial to students, most felt unprepared to lead standards-driven classrooms. Reporting the comments of a teacher at one school site, an observer said that this teacher believed she needed a lot more professional development before she could have a standards-driven classroom. This teacher said, "I'm not even really sure what that means." Unfortunately, there was little guidance from School District personnel to help her define it.

Finding: The Performance Responsibility Index (PRI)—the new accountability system—was not well understood or accepted during the 1996-97 school year. However, teachers did seem to believe that it had the potential to contribute to improved system performance.

On the survey, 51 percent of teachers reported that they understood the purpose of the PRI. The interview data clearly illustrated that most teachers were apprehensive and even resentful of the new performance system. Many teachers seemed to think of it as a punishment specifically aimed at them. As one teacher explained:

Teachers are feeling very resentful about what is happening in the system. Many new teachers are thinking of leaving because there is no support. These kids are challenging. We don't need to pick on teachers and say they are not trying. The administration is placing blame for system failure on the teachers. It's undermining us.

The PFT bolstered this view, arguing that the procedures and sanctions in the PRI were focused on teachers and were designed to serve them up as scapegoats. Teachers expressed concern that middle level administrators and principals were not being subjected to the same level of oversight and threat of dismissal or the denial of wage increases. The teacher survey also indicated that most (88 percent) were afraid that the PRI would unfairly reward and punish many schools.

In addition to resentment, when the scores from the first round of the PRI were publicized many teachers felt humiliated and betrayed. One said, "They told us the first administration of the test was baseline, that it didn't count, but then Hornbeck broadcast the scores and everyone thinks we're bad." This sentiment was widespread among the teachers interviewed, and it was reflected in the June 1997 teacher survey—at that time only 15 percent of teachers in the District believed they were respected by Superintendent Hornbeck.

Despite the mostly negative reaction to the implementation of the accountability system, some teachers did believe that it might help their students. Of the 51 percent of teachers who reported understanding the purpose of the PRI on the survey, 63 percent of them believed that it had the potential to benefit their students. This suggests that if District officials make a greater effort to educate teachers about the purpose of the PRI, more may see its potential benefit. In addition, 42 percent of teachers believed that the new responsibility system will cause teachers to increase efforts to improve teaching. This finding indicates that while the punishment aspect of the accountability system may not be welcome, it may serve as an incentive for teachers to take a critical look at how they can improve student achievement.

Finding: Many teachers feel that they are being held accountable for achieving results that are beyond their control.

The School District of Philadelphia content standards serve at least two functions. One function is to promote better teaching practices by making clear what it is that Philadelphia children should know and be able to do.[6] Another is to hold teachers accountable for teaching and for students learning particular subject matter and skills. This latter purpose of standards is particularly problematic in the District because many teachers feel that numerable factors beyond their control limit their teaching effectiveness and their students' ability to meet the standards. A majority of Philadelphia teachers seemed to question whether the city's school children were capable of meeting the standards. Table 6 below shows that over 59 percent of high school English teachers felt that fewer than half of their students were capable of reaching the English standards; 71 percent of high school mathematics teachers felt that fewer than half of their students were capable of reaching the math standards; and over 75 percent of high school science teachers felt that fewer than half of their students were capable of reaching the science standards.

[6] See Simon, E., 1998

TABLE 6

Teachers' Beliefs about Students' Abilities to Meet the Standards in Various Subjects and Obstacles to Achieving Standards*

Survey of Philadelphia Teachers, 1997

Item	Percent of Teachers Overall	Percent of High School Teachers
About how many of your students would be able to meet the standards in...	**Fewer than half to none**	**Fewer than half to none**
English and Language Arts	53.5 N=957	59.4 N=96
Mathematics	51.9 N=883	70.7 N=82
Science	55.4 N=832	75.4 N=69
How important are each of the following factors in hindering your students' success in achieving the standards?	**Important or moderately important**	**Important or moderately important**
Lack of basic skills	96.6 N=1,303	96.4 N=220
Inadequate prior student preparation in the subject area	92.9 N=1,292	94.5 N=205
Inadequate motivation for education among students	94.0 N=1,290	95.0 N=220
Inadequate instructional materials	85.5 N=1,288	82.6 N=218
Lack of teachers' mastery of content area	73.7 N=1,263	70.5 N=207
Inadequate alignment of curriculum & standards	81.9 N=1,260	81.1 N=212
Lack of teacher consensus on standards' appropriateness	73.1 N=1,251	65.9 N=205
Lack of teacher skill in communicating content	79.0 N=1,251	71.9 N=210
Inadequate additional support for students who need it	93.4 N=1,272	92.1 N=214
Inadequate subject matter articulation across grade levels	84.0 N=1,244	78.7 N=21
Student advancement to grade without meeting promotion requirements	94.5 N=1,266	93.9 N=212
High student mobility in and out of the school	88.7 N=1,267	89.9 N=215
Poor student attendance	92.4 N=1,291	98.2 N=219

32

*Note: Excludes teachers working in special admission schools.

Research has shown that many urban teachers hold low expectations for the students they teach (Abelman, Elmore & Kenyon, 1997), and additional data from Table 6 suggest that many Philadelphia teachers share that characteristic, particularly in regard to standards. Over 90 percent of teachers reported that factors like lack of basic skills, poor attendance, and inadequate motivation for education were important or moderately important factors hindering students from achieving the standards. Teachers saw these factors as more important than teacher actions or school policies and conditions.

All of this data was supported in qualitative interviews with teachers. They consistently described students who lacked motivation for education and whose attendance was poor. A first-year high school teacher told us, "The attendance rates are atrocious. You have to sit on top of [students] to do work." A few also mentioned the difficult circumstances many of their students face, including poverty, homelessness, and, most commonly, unsupportive, negligent or abusive families. Lack of parental support was a consistent theme. To help their children achieve academically, "parents have to learn how to be parents," and have "to assume their rightful responsibilities," teachers told us.

A white teacher at a predominately minority high school responded vehemently to the District's adoption of the Stanford-9 Achievement Test (SAT-9), noting particularly that responding in short essays was beyond his students. He felt it was unfair to administer the SAT-9 to "these students Most of these students were not equipped to deal with it. It is beyond most of them. It's not appropriate for these students." When we asked him how he prepared them for it, he noted that he emphasized writing and structuring paragraphs. He said though that "trying to squeeze a five sentence paragraph out of these kids is like [asking them] to put their head in a juicer."

To some degree these attitudes stem from the racism and classism that permeate our society and are reproduced in schools (Anyon, 1997), especially in a system where over 60% of teachers are white, but 80% of the students are racial or ethnic minorities (School District of Philadelphia, 1996). But it is also important to note that experience and realism may also be factors. Many of Philadelphia's public school students do face a number of circumstances that challenge their ability to succeed in school, and many teachers do find the task of teaching those students daunting.[7]

Whatever its source, this skepticism about the ability of Philadelphia children to reach the standards led many teachers to either dismiss the new standards, assessment and accountability system as unrealistic or lash out against it as unfair to them. On the survey, 71 percent of teachers reported that they believed that their success or failure in teaching was due to factors outside of their control. One teacher summed up the sentiments of many when she said:

It holds us accountable for things we have no control over. I can't control whether a kid comes to school. You can hold me accountable for the kids who attend and are in class and have the materials. I'll take accountability for these, but why hold me accountable for the kid who comes once a month?

The survey results certainly indicate that teachers believed that whether or not students succeeded was largely dependent on factors outside of the school system. And this raises questions about who should really be held accountable for student results. After all, students play the central role in their own education. They must be motivated to do the work necessary to acquire understanding. Although teachers play a role in whether or not students succeed, they are often frustrated by what they can accomplish on their own. As one teacher said, "We are tired of being held totally responsible for someone else's job. We aren't here to become nurses, social workers, or moms and dads. We are trained to be teachers and we can't do our jobs because of other factors."

[7] For more information on the life circumstances of Philadelphia children, see Foley, E., *Restructuring Student Supports: Redefining the Role of the School District.*

Finding: The implementation of the Stanford-9 Achievement Test impacted teaching practice.

On the survey, 80 percent of teachers said they agreed that it is a good idea to have a District-wide measure of student performance, and 73 percent of teachers reported they understood the purpose of the SAT-9. Of those teachers, 60 percent also reported that it had the potential to benefit their students. Some teachers felt that the test encouraged critical thinking and problem solving and that it led to better teaching. As one teacher reported, "The SAT-9 is a good idea; it asks more critical thinking problems, and it pushes me to do that kind of thinking with my class." The assessment helped some teachers realize they needed to move away from rote learning and have kids do more performance tasks, critical thinking, writing and problem solving[8].

In addition, the assessment alerted school staff to the areas in which they need to make significant improvement. A number of principals noted that the results were used to choose areas to highlight in their school improvement plans. One teacher said, "The SAT-9 showed us that the students comprehension skills aren't where they should be or report card marks in general." Another teacher said that the staff at her school "looked at the SAT-9 test scores to develop strategies to create curriculum to improve test scores. The student body is great and the staff is great but the scores are not acceptable, so we are making drastic changes in our math program."

Since the SAT-9 scores count for 60 percent of the total on the Performance Responsibility Index, school staff did everything they could to improve student achievement on the test. Most (73 percent) also reported that they felt pressured to improve student test scores. As a result, an almost universal response to the implementation of the SAT-9 was to take steps to prepare students for the test. As one principal explained, "The professional responsibility system has raised people's interest in the SAT-9 and the school bought Key Links booklets to help the kids get ready for the test." Since fieldwork was conducted close to the time of the SAT-9 test administration, researchers observed—in nearly every school—test prep classes. These classes were designed to expose students to the types of questions asked on the SAT-9. In some cases, teachers believed this form of "teaching to the test " was generally beneficial and hoped to integrate some of the strategies into their own curriculum. However, in other cases, teachers saw getting students ready for the test as a waste of time and an interruption. On the teacher survey, 36 percent of respondents reported that they believed too many teachers spent too much time on test-taking skills in preparing for the SAT-9. Other teachers thought it was ridiculous that they had to put their regular curriculum on hold for the weeks surrounding administration of the test.

In addition to preparing students, a tremendous amount of time was spent on preparing and familiarizing teachers with the SAT-9 test. In the qualitative research, nearly every teacher reported attending some type of professional development session on the SAT-9. Most of this professional development focused on effective instructional strategies for problem-solving and critical thinking skills. This effort to familiarize teachers with the assessment seemed to pay off. On the survey, 71 percent of teachers reported being familiar with the content of the SAT-9.

Finding: Few teachers believed the SAT-9 reflected the new standards or their curriculum.

Although the data presented in the preceding section clearly demonstrates that most teachers saw some benefit from the administration of the SAT-9, only a minority reported on the survey that they believed it was a good measure of the content standards. In addition, only 27 percent believed that the test was in line with the subject matter they taught. On the survey, almost 81 percent of teachers reported that con-

[8] For more information on how teaching practice has changed, see Simon, E., *Making Sense of Standards*.

tent standards had the potential to benefit their students, and 80 percent also agreed it was a good idea to have a District-wide measure of student performance. However, only 35 percent reported that they believed the SAT-9 was well aligned with the content standards. What explains this discrepancy in how teachers felt? One cause may be the media campaign waged by the Philadelphia Federation of Teachers (PFT) against both the standards and the assessment. Federation officials saw the content standards as too vague and they have argued through the media and through their own direct mail to teachers that Philadelphia needed "real education standards and a clear curriculum with subject matter content." Since teachers reported that they did not have enough to time to fully implement the standards or prepare their students for the assessment, it is doubtful that many took the time to individually match up the standards and the assessment. So their views about the SAT-9's alignment with the standards may well have reflected what they heard from union officials.

Even though the SAT-9 was well received by most teachers, on the survey only 27 percent of teachers reported that the SAT-9 was well aligned with the subject matter and the grade level they taught. The disjuncture between how positively teachers felt about the SAT-9 and whether or not they believed it was aligned with their curriculum may be a consequence of where teachers believe students are developmentally. While many felt that the SAT-9 was a good test abstractly, they may not have believed it was appropriate for their own students for various reasons. As one teacher said, "The students aren't being tested on what they were taught, so why are we using this? We need a test appropriate for our own students." A lot of teachers worried that the test was simply "too hard" for their students. One of the teachers who served as a proctor during the test administration noted, "The students did not want to take it because they did not see the value. They looked at it and couldn't do it, so they gave up." Many teachers noted that their students were "shocked" by the test and complained that they had never seen most of the material before and were not used to writing out answers. One reason teachers did not believe their students could do well on the test was because they believed that their students did not come in to their class with the necessary prerequisite knowledge and skill level. On the survey, 51 percent of teachers agreed with the statement that students did not have the prerequisite knowledge required to do well in their class. In addition, over 60 percent of teachers reported having to spend three or more weeks reviewing content that their students "should already know."

TABLE 7

Time Spent on Review

Survey of Philadelphia Teachers, 1997

Question: In your target class, about how much of your teaching this school year has been spent teaching or reviewing content and skills you expected students to have learned at previous grade levels?	**Percentage**
None	10.6
1 week	6.2
2 weeks	16.8
3 to 5 weeks	30.8
6 to 9 weeks	14.1
10 or more weeks	21.4

Finding: The new teacher evaluation form was not well understood.

As discussed in the section Elements of the *Children Achieving* Accountability System, the District created a new observation form for principals to use in evaluating teachers to ensure that teachers were implementing the new standards in their classrooms. In most schools, both teachers and principals seemed unaware of the changes in the form, or if they knew of the changes, simply believed that "it's really not that different than the old form." In many cases, teachers said it had been so long since someone observed their classroom that they did not know anything about any of the evaluation forms.

However, the new rating form did result in strong concerns in a few schools. One principal said that when he introduced the new observation form in his school, a rumor started among the teachers that the new rating system was specifically designed to get rid of people because of budget cuts. Several teachers said they believed a lot more of their colleagues were being rated "unsatisfactory" because of the new form even though "these are good teachers." At one school there seemed to be a widespread belief that teachers could be coded unsatisfactory for such trivial things as having the blinds on their classroom windows uneven.

On the other hand, the new observation form was credited with having positive impacts as well. A principal noted that the teachers in her school did not believe in the standards, did not take them seriously because they thought they were a passing fancy until the new teacher observation form was released. Because the observation form requires the principal to determine if the appropriate standards are being covered in the observed lesson, this "let teachers know that they couldn't blow the standards off."

The RESULTS *of the* FIRST ACCOUNTABILITY CYCLE

The tables on the following pages give an overview of the SAT-9 and PRI results. The size and scope of the gains are illustrated by level, by subject, and by first and second cohort clusters.

TABLE 8

School District of Philadelphia 1996-97 Student Achievement Data

Stanford-9 Achievement Test Results, by Subject and Cluster: Elementary, K to 8 and Middle Schools

CLUSTER	Promotion Rate	Percentage of students scoring at or above basic on the SAT-9 in READING						Percentage of students scoring at or above basic on the SAT-9 in MATHEMATICS						Percentage of students scoring at or above basic on the SAT-9 in SCIENCE					
(excluding Special Admit Schools)	Grades Other	ELEM	Change since '96	K-8	Change since '96	MIDL	Change since '96	ELEM	Change since '96	K-8	Change since '96	MIDL	Change since '96	ELEM	Change since '96	K-8	Change since '96	MIDL	Change since '96
Audenried	83.7	38.8	6.8	N/A	N/A	40.6	5.1	32.3	10.2	N/A	N/A	14.8	6.4	38.1	9.2	N/A	N/A	18.0	6.7
Bartram	87.7	45.4	7.6	N/A	N/A	52.3	7.5	41.1	11.5	N/A	N/A	17.4	1.9	48.9	13.7	N/A	N/A	23.1	5.6
CHAIN	92.8	72.3	9.4	87.9	11.2	71.2	12.9	70.6	13.3	77.9	13.8	43.8	8.1	72.6	11.7	86.4	19.4	48.3	10.2
Edison	79.9	47.0	13.9	N/A	N/A	29.3	1.3	41.7	14.7	N/A	N/A	3.1	-1.3	41.4	13.9	N/A	N/A	9.9	5.1
Fels	88.4	51.7	-0.5	65.1	3.9	N/A	N/A	47.9	4.8	44.7	0.8	N/A	N/A	50.2	5.7	47.6	4.1	N/A	N/A
Frankford	89.6	60.7	6.4	65.7	5.2	59.3	19.0	56.5	10.8	45.6	11.5	21.1	1.5	58.9	8.4	51.8	15.2	28.0	5.8
Franklin	88.6	44.7	15.9	63.7	-6.4	40.8	14.6	38.7	10.9	48.8	4.2	5.5	3.9	40.5	11.2	50.0	-0.8	12.7	6.7
Furness	91.2	N/A	N/A	63.3	5.0	N/A	N/A	N/A	N/A	54.2	13.0	N/A	N/A	N/A	N/A	55.8	14.3	N/A	N/A
Germantown	88.6	46.4	9.6	78.8	7.2	46.4	9.9	34.8	6.0	55.2	8.3	6.7	1.0	40.6	7.2	60.5	8.0	14.9	5.9
Gratz	86.0	41.0	8.3	N/A	N/A	38.0	8.3	36.4	12.4	N/A	N/A	12.8	6.8	36.8	6.8	N/A	N/A	14.4	7.5
Kensington	82.8	47.5	4.0	54.7	11.5	43.0	7.2	44.8	7.1	45.9	11.0	11.9	-1.2	47.8	9.6	43.4	10.3	14.7	0.0
Lincoln	87.2	68.0	4.9	70.1	4.4	66.1	8.4	70.1	14.9	44.9	1.6	29.5	4.5	75.8	13.6	57.5	10.2	35.2	2.5
ML King	87.4	43.9	2.6	N/A	N/A	49.9	2.9	34.5	4.1	N/A	N/A	21.8	13.0	41.3	3.8	N/A	N/A	23.7	10.7
Northeast	90.2	77.6	7.5	84.8	9.6	69.5	2.3	69.0	4.4	72.0	16.6	29.3	-3.7	71.3	5.8	74.0	14.7	43.7	7.5
Olney	85.0	44.8	9.2	66.7	8.1	45.3	5.9	38.2	10.4	59.0	20.1	11.2	2.8	42.9	12.9	58.0	18.5	16.0	4.0
Overbrook	87.7	50.3	12.5	N/A	N/A	42.5	-4.2	35.3	6.2	N/A	N/A	15.1	-0.2	42.9	5.9	N/A	N/A	18.6	5.2
Roxborough	88.4	45.9	0.9	66.2	8.5	N/A	N/A	36.7	14.2	44.1	3.5	N/A	N/A	37.8	-11.0	50.5	8.0	N/A	N/A
South Philadelphia	88.1	64.5	9.1	59.4	-4.5	51.0	8.9	53.6	10.0	42.4	-1.2	16.9	3.3	63.8	15.4	42.9	-1.1	19.8	3.4
Strawberry Mansion	84.3	39.6	4.5	N/A	N/A	48.1	15.7	34.2	5.3	N/A	N/A	5.8	-1.1	46.4	14.9	N/A	N/A	13.1	1.1
University City	86.0	47.1	10.1	57.3	6.4	38.8	16.9	37.3	9.2	18.1	7.3	5.6	-0.7	42.2	6.1	24.4	9.3	8.2	2.2
West Philadelphia	86.6	43.7	8.5	47.3	0.6	41.5	5.0	34.2	12.4	36.1	7.2	8.4	-2.7	38.1	6.3	38.8	6.4	13.5	2.8
William Penn	84.4	37.4	17.2	N/A	N/A	33.0	1.4	30.6	19.6	N/A	N/A	7.5	5.2	31.8	14.5	N/A	N/A	8.8	5.6
Special Admission	89.1	54.3	8.9	78.2	19.5	87.7	5.0	21.7	0.5	67.9	18.7	51.8	-4.8	39.1	5.8	69.2	10.5	62.9	5.4
District Total	**87.0**	**48.7**	**8.2**	**64.9**	**5.2**	**50.5**	**7.1**	**41.9**	**9.7**	**48.4**	**8.0**	**18.5**	**2.8**	**46.4**	**9.3**	**51.8**	**9.6**	**23.7**	**5.6**

BEST COPY AVAILABLE

TABLE 9

School District of Philadelphia 1996-97 Student Achievement Data
Stanford-9 Achievement Test Results, by Subject and Feeder Cluster: High Schools

			Reading		Mathematic		Science	
CLUSTER	**School Name**	Persistence Rate Grades 9-12	% at or Above Basic	Change since 1996	% at or Above Basic	Change since 1996	% at or Above Basic	Change since 1996
Audenried	Audenried High School	19.3	6.1	3.9	1.0	1.0	0.0	0.0
Bartram	Bartram, J. High School	42.7	30.8	16.1	7.1	4.9	4.1	3.4
CHAIN	Washington, G. High School	63.6	40.5	7.2	23.1	-0.6	15.9	6.1
Edison	Edison High School	24.4	12.4	4.2	2.6	1.6	0.8	0.8
Fels	Fels High School	41.3	37.7	14.6	18.5	10.8	8.0	4.6
Frankford	Frankford High School	35.6	29.7	8.1	5.9	-0.3	2.4	.3
Franklin	Franklin, B. High School	34.4	14.0	6.5	2.9	2.2	1.0	1.0
Furness	Furness High School	32.6	25.4	12.9	13.5	8.8	4.1	3.6
Germantown	Germantown High School	35.1	26.2	-3.4	2.9	-2.2	1.5	-0.3
Gratz	Gratz High School	20.4	9.9	5.0	0.6	0.3	0.0	-0.3
Kensington	Kensington High School	16.9	15.7	10.1	3.1	1.3	1.9	1.3
Lincoln	Lincoln High School	44.8	37.0	11.9	11.1	6.1	5.9	3.6
ML King	King, M. L. High School	39.1	14.3	-4.2	2.9	1.4	0.4	0.4
Northeast	Northeast High School	61.9	44.4	18.4	20.6	6.8	12.3	4.3
Olney	Olney High School	25.0	16.7	11.9	8.4	6.9	3.2	3.2
Overbrook	Overbrook High School	43.9	25.5	20.2	1.6	1.6	0.8	0.8
	Lamberton School	52.2	68.1	19.8	27.9	2.3	35.9	5.0
Roxborough	Roxborough High School	52.1	38.9	22.1	10.2	9.2	4.9	4.2
South Philadelphia	South Philadelphia High School	26.1	18.1	5.0	5.8	-0.6	2.3	2.3
	Girard/GAMP	93.5	71.8	6.8	56.4	11.8	55.7	11.5
Strawberry Mansion	Strawberry Mansion School	28.9	22.5	-5.6	5.2	-2.7	4.6	0.7
University City	University City High School	33.2	18.7	5.7	3.3	-1.0	0.5	0.0
West Philadelphia	West Philadelphia High School	28.8	12.9	3.3	1.6	0.8	0.0	0.0
William Penn	Penn, Wm. High School	31.2	18.3	10.8	1.5	1.5	1.0	1.0
Special Admission	CAPA High School	85.8	68.7	15.7	18.4	8.2	14.7	10.5
	Bok AVT	61.8	13.5	-2.2	2.2	-0.3	1.1	1.1
	Masterman School	98.9	98.9	0.4	95.8	0.9	90.5	7.6
	Franklin Learning Center	63.5	67.4	43.6	20.6	16.4	6.9	4.3
	Carver High School	83.3	89.3	0.8	68.4	11.5	33.3	10.9
	Dobbins AVT	56.3	24.9	14.8	1.4	0.9	1.6	1.1
	Mastbaum AVT	48.6	30.7	-0.6	6.9	2.7	3.2	2.0
	Parkway Programs	70.2	58.7	13.5	8.1	5.1	2.0	-0.2
	Bodine High School	87.5	77.9	9.1	25.7	6.9	15.4	1.3
	Central High School	91.8	92.8	7.3	80.4	3.2	56.3	9.9
	Saul High School	73.1	61.0	14.1	19.5	4.9	13.4	-0.4
	Girls' High School	90.8	88.7	1.0	54.6	-1.9	26.8	20.2
Subtotal	Special Admission Schools	64.2	59.3	6.9	34.4	1.9	23.7	5.9
Total	**District High Schools**	**43.4**	**37.1**	**8.7**	**16.9**	**2.7**	**11.7**	**3.6**

BEST COPY AVAILABLE

TABLE 10

School District of Philadelphia 1996-97 Performance Responsibility Index Data
PRI Scores, by Cluster

Cluster	1997 PRI Score	Change in PRI Since 1996	1997 Below Basic Reading	1997 Below Basic Math	1997 Below Basic Science	1997 Average Percent Below Basic, All Subjects	Decrease in Average Percent Below Basic Since 1996
Audenried	**63.0**	**6.6**	**62.7**	**76.3**	**71.5**	**70.2**	**7.5**
Bartram	66.2	8.5	55.9	75.2	70.7	67.3	8.9
CHAIN	**73.6**	**7.4**	**36.9**	**53.0**	**52.0**	**47.3**	**8.9**
Edison	59.6	7.5	63.4	76.9	75.5	71.9	7.3
Fels	68.2	5.7	45.8	64.3	64.8	58.3	4.5
Frankford	68.8	5.5	45.5	65.2	62.2	57.6	7.2
Franklin	66.5	8.5	50.7	66.9	67.2	61.6	9.6
Furness	69.1	5.9	46.1	60.3	62.3	56.2	8.0
Germantown	69.0	5.2	43.5	67.9	66.3	59.2	5.8
Gratz	58.5	6.4	66.5	79.9	79.4	75.3	7.4
Kensington	65.0	6.2	51.0	67.2	64.6	60.9	3.9
King	**66.6**	**3.5**	**49.3**	**65.2**	**66.9**	**60.5**	**4.8**
Lincoln	68.2	5.7	42.6	65.4	61.9	56.6	4.5
Northeast	74.3	7.9	34.1	55.7	54.5	48.1	7.6
Olney	**61.7**	**6.5**	**57.6**	**73.9**	**72.3**	**67.9**	**8.5**
Overbrook	63.7	6.8	54.1	77.4	72.8	68.1	6.8
Roxborough	67.8	8.0	43.5	70.1	68.1	60.6	8.0
South Philadelphia	65.8	3.9	50.0	67.6	66.1	61.2	4.3
Strawberry Mansion	**59.2**	**1.2**	**65.4**	**82.3**	**76.1**	**74.6**	**2.8**
University City	60.1	6.4	59.2	82.3	80.0	73.8	6.4
West Philadelphia	**62.1**	**4.7**	**60.8**	**77.5**	**74.3**	**70.9**	**5.0**
William Penn	60.0	9.1	58.6	78.1	82.9	73.2	8.5
Primary (K-8) Total	**70.3**	**5.7**	**47.0**	**64.7**	**60.4**	**57.4**	**7.1**
Secondary (9-12) Total	**48.3**	**7.5**	**65.2**	**85.3**	**91.4**	**80.6**	**5.0**
District Total	***65.7***	***6.1***	***51.8***	***70.1***	***68.5***	***63.5***	***6.6***

First Cohort Clusters

BEST COPY AVAILABLE

Controversy over the Gains

As the preceding tables clearly demonstrate, between the 1995-96 school year and the 1996-97 school year, student achievement on the Stanford-9 Achievement Test increased significantly, which led to increases in the overall school scores on the PRI. Ninety-two schools had more than a five point gain at the basic level or above, and over one-half of those had more than a ten point gain. The Superintendent and others viewed these improvements as a major achievement and as affirmation that the *Children Achieving* reform agenda is a sound one. The gains, however, were not without controversy.

Soon after the 1996-97 scores were released, some School Board members and people in the press questioned the legitimacy of the increases. They variously argued that the scores increased simply because more students took the test, or because the baseline was very low, or because of the methods used to calculate the index. These questions were raised publicly when the Superintendent met with the School Board in January. The questions reflect misunderstandings of the purpose and the structure of the index. The following section describes how the index is actually calculated and the issues associated with it.

Calculation of the Index

Stanford-9 Achievement Test. As noted earlier, the three SAT-9 tests in reading, mathematics and science comprise more than 60 percent of the performance index score. We observed educators in many schools making focused efforts to improve their students' achievement on these tests. Yet there were questions about how these scores were used in the PRI.

A primary controversy concerned the number of students considered "not tested." In order to encourage the testing of all students, regardless of ability, the School District made a policy decision to count all students who were "not tested" in each subject area as zeros in calculating the overall score for a school. This prevents schools from attempting to raise their performance on the SAT-9 by pushing low performers out. Yet, the "not tested" category is a misnomer. It implies that all of the students in that category were either absent or not included in the testing. Many of the students who were counted as not tested did participate in the examinations, but either they simply did not complete all of the sub-tests for that subject area (e.g. they did not take one of the two reading sub-tests) or for various reasons they did not receive a valid score on the test. The exact criteria applied varies by item type, but generally, to receive a valid composite score in each subject test, a child must:

- attempt either three out of the first six items, or any ten multiple choice questions AND get one correct answer; and
- attempt one open-ended question and be credited at least one point by the scorer.

Some educators believed that this policy unnecessarily penalized schools with large special education populations who they hypothesized had a higher "tested, but invalid" rate than other schools. We heard many teachers tell us about the frustration experienced by their students when the tests were first administered in the baseline year. However, other educators believe special education students are penalized if the system does nothing to encourage their inclusion.

Teacher and student attendance. Calculations for measuring teacher and student attendance are perhaps even more complex than the scoring of the SAT-9. Because of this complexity and concerns about the calculations, the attendance indicators are analyzed more closely in this section. We examine first the calculation of the student attendance component of the PRI.

Student attendance is not measured using a traditional average daily attendance score which would average over a specific time period the percent of students attending school each day out of the total number enrolled daily. Instead, the School District utilizes a complex formula which is based on the total number of student days for each school. All students enrolled at any point during the school year are included in this statistic by measuring the percentage of days each student attended out of all the days he or she was enrolled at the school. The measure is calculated in this manner in order to hold schools accountable for **all** students, not just those students who were enrolled for the entire year. The total number of student days is calculated by:

1. Calculating the total number of students enrolled in the school at any point during the school year (including those that, for example, attended for a school for four weeks and then transferred to another school; or those that enrolled in the middle of the year);
2. Calculating the total number of days it was possible for each student to attend;
3. Summing the possible days for each student to arrive at the total number of possible student attendance days.

To then calculate the actual student attendance score:

1. Each student's individual attendance rate is measured and the students' days enrolled are placed in one of the following categories: Advanced (96-100%); Proficient (95%); Basic (85-94%); Below Basic III (80-84%); Below Basic II (75-79%); and Below Basic I (10-74%).
2. Then the total number of days enrolled by students in each category is summed and divided by the schools' total number of possible days to arrive at a PRI score for each category. For example, if the total possible days for a school was 30,000 and the number of days enrolled by students in the advanced category was 3,000, students in the advanced attendance level would account for 10 percent of the total possible days for the school.
3. Attendance rates in the advanced category are then weighted by a factor of 1.2, so the total sub-score in the advanced category would be 12.
4. This is then repeated for each category (with different weights depending on the scoring category) and summed to come up with the total student attendance score.

Staff attendance rates are calculated using an easier method. They are based on the total number of staff and the percent of school days they attended. All teachers, as well as administrative & instructional support staff, are included in this statistic. Food service workers and custodians are not included in the score. The same weighting system is used by category (i.e. advanced is weighted 1.2; proficient is weighted by 1.0; basic is weighted 0.8, etc.), but the cut scores are different. For example, any staff attendance under 93% is considered below basic, while student attendance under 85% is considered below basic. Additionally, the differences between basic and advanced attendance are so small that a teacher who attended 169 days out of a 180 day school year, one who attended 171 days, and one who attended 173 days would each be included in a separate category. Additionally, a school in which 60

percent of staff attended 96% of school days would receive 72 points toward their overall staff attendance score, while the same proportion of teachers attending 94% of school days would receive only 48 points toward their overall score. These varying cut scores and complexities of calculation raise questions about how these cut scores were developed, their validity, and the rationale for applying the SAT-9 categories of Advanced through Below Basic to indicators like staff and student attendance.

Another controversial aspect of the staff attendance score involves who is included in the calculation. As noted above, some school support staff such as food service workers and custodians are excluded. But teaching staff on long-term leave for illness or maternity leave are not excluded. This has struck many principals and staff as unfair and illogical.

Finally, one last complexity of the attendance scores is how they contribute to the overall calculation of the Performance Index. While the SAT-9 results for each subject area count individually as one component of the total PRI score, staff and student attendance are averaged to create what the District calls an "enabling score." These two different indicators were collapsed into one component to make one composite score which counts for one-fifth of the total PRI score.

It should be noted that District officials chose to include student and staff attendance as a variable in the index because chronic absenteeism is a problem in Philadelphia schools. On any given day at many of the city's comprehensive high schools, only half of the enrolled students attend class all day. Teacher absenteeism is also a concern. Even in comparison to other urban school districts, which suffer from high rates of teacher absenteeism, Philadelphia has poor teacher attendance.

Promotion/persistence. Promotion and persistence rates are calculated differently than any other indicator in the PRI. The promotion rate measures the percent of students who met the requirements of one grade and were passed on to the next. It is used exclusively in elementary and middle schools. In high schools, promotion rates are not calculated as a student has to receive a certain number of credits to graduate, and does not necessarily move from grade to grade in the same systematic fashion as an elementary curriculum would promote. High school persistence rates measure the percent of students who enter the school in the ninth grade and then graduate on time from that school or any other Philadelphia high school four years later.

Unlike the other components of the PRI, the promotion/persistence rate is a school statistic, not an individual teacher or student characteristic. While test scores and teacher and student attendance scores yield a series of sub-scores for each category, promotion and persistence rates yield one score per school. That is, if in one elementary school 96% of students are promoted, the school receives an advanced score for this component of the index. However, also unlike the other components of the PRI, promotion/persistence is an unweighted score. For example, if an elementary school has a promotion rate of 96.2, it receives a score of 96.2 in the persistence component of the PRI. As noted above, on the testing and attendance components, advanced and basic and below basic scores are weighted to be higher or lower according to performance on the indicators. This is the result of concerns by District staff that weighting promotion rates would encourage schools to simply advance more students to the next grade in order to raise their index score.

In some schools (for example, K-12 schools), the promotion and persistence variables are combined, which has caused concern. Promotion and persistence are very different variables. Many high schools feel they are unfairly penalized in comparison with elementary schools. They argue that elementary schools can promote students who have not met the requirements of their grade, and then the high schools are blamed for failure when students cannot graduate on time.

External Review of the System

As a result of the questions about the PRI and the SAT-9 raised by the school board, teachers, and the media and concern about some statistical errors made by Harcourt Brace, the test publisher, the Superintendent agreed in January 1998 to have outside experts evaluate the PRI. Researchers from CPRE and RAND have assisted with the development of a review process. The plan is to assemble a panel of experts to look at issues around design, administration and behavioral response, reliability, validity, and impact. Examples of the types of questions the panelists might consider are outlined below.

Design Issues

- Are the performance standards set at reasonable levels?
- How large and rapid are the gains schools are expected to make and are these reasonable (in terms of rate as well as magnitude) in light of past experience and research?
- What incentives does the system establish for long-term vs. short-term change?
- Is the SAT-9 an appropriate assessment for this high-stakes use? Is it sufficiently secure?
- Are the indicators included in the PRI appropriate?

Reliability

- What is the error associated with classifications of schools? Which schools are most affected?

Validity

- How adequate is the curricular basis for the SAT-9? Does it match the standards, and is it mapped to a clear set of curricular expectations?
- Are the scale and reporting metric reasonable and robust?
- What mechanisms are in place to monitor teaching to the test and inflated test scores or to lesson the likelihood of Lake Wobegon effects?

Impact

- What evidence is being collected about the effects of the Professional Responsibility Index on school organization, the implemented curriculum, training and various aspects of instructional quality?
- Is the Professional Responsibility Index sufficiently understandable to administrators, teachers, and parents so that they can respond appropriately?

Identifying Failing Schools

As noted in the previous section, most of Philadelphia's 259 schools showed some improvement on the Performance Responsibility Index between 1996 and 1997. However, 15 schools were identified as needing assistance because their scores declined. A few of these schools had relatively high overall performance on the SAT-9 but because their PRI score declined compared to the previous year, the support process was triggered automatically. This is because the objective of the accountability system is not to compare schools to each other, but to emphasize a process of continuous improvement and help all students achieve more, not fall further behind.

As Table 11 below demonstrates, Conwell (a special admissions school) outperformed Edison High School on every measure. However, because Conwell declined on the PRI between 1996 and 1997, it has been labeled a "low progress school."

TABLE 11

Performance Comparison of Two Schools

	PRI Score 1996	PRI Score 1997	PRI Point Change	PRI 1998 Target	Below Basic 1996	Below Basic 1997
Conwell	88.5	87.6	-.9	89.6	13.1	19.7
Edison H S	27.2	33.8	6.6	38.5	97	94.8

To assist the 15 schools labeled low progress, the District assigned each one a "school support team." These teams are chaired by a District official but include parents, school staff members, and teachers and administrators from outside the school. In October 1997 these teams conducted observations at the schools, met with parents and staff, and developed recommendations for each site. The 15 schools are expected to implement the recommendations within a specified time frame. Progress in improving student achievement will be monitored by the District, and if the improvement plans prove to be ineffective, these schools could be eligible for reconstitution in 1999. Although the low progress schools receive recommendations and support from the central office, they receive no additional monetary assistance.

The table above raises the question, what does it mean to be a low progress school? Should a school, such as Conwell, that is only 7.4 points away from obtaining the 12-year district goal of a score of 95 on the PRI be labeled "low progress"? Is it appropriate to target additional resources and assistance to a school like Conwell when there are numerous schools like Edison where almost 95 percent of the students are scoring below the basic level in reading, math, and science? This point is not lost on District officials or the Superintendent. They argue that the goal of the PRI is to create a level playing field, so schools are not competing with each other and should not be compared to each other. The ultimate goal is for all schools, regardless of achievement level, to move further down the path of continuous improvement. In this view, equity demands that the District work to improve all schools in the system whether they are close to achieving the goals set out in the PRI or have a long road ahead.

Controversy Over Keystoning

As mentioned earlier in this report, the Superintendent has the authority to reconstitute schools, which allows District officials to force a transfer of up to 75 percent of the staff. This authority was granted to the Superintendent in February 1995 through an agreement with the Philadelphia Federation of Teachers that is subject to numerous restrictions. However, when the PRI was developed, reconstitution was added as a sanction in that process. As outlined in the PRI, schools at risk of reconstitution are those that fail to meet their short-term goals for two rating periods (four years) in a row. Although schools had only gone through one rating cycle by February 1997, the Superintendent officially announced that two high schools, Olney and Audenried, would be designated as Keystone schools and reconstituted. The decision for selecting the schools for reconstitution was based on test scores, graduation rates, and student and staff attendance rates, adjusted for the level of poverty at the school. Additionally, the Superintendent asserted there was an "inability of the school staffs to work as a team" to improve the schools.

What ignited a firestorm of controversy was the way the reconstitution plan was announced and introduced at the schools. It came as a surprise to almost everyone. Although the Superintendent had the contractual authority to reconstitute the schools, they had not yet gone through the two annual ratings that constitute a cycle in the accountability system. However, Superintendent Hornbeck argued that the criteria for reconstitution were clear before the decision was announced. The issues were performance and whether there was a team at work in the school that gave him reason to believe that performance could be raised. District officials argued that Audenreid had already been through one school review process, had more monetary support because it was one of the first cohort clusters, and was still one of the lowest performing schools. In addition, Olney also had first cohort resources, a new principal, and was performing below other schools of a similar poverty level. Despite this rationale, soon after Hornbeck announced his decision to reconstitute the two schools, the Philadelphia Federation of Teachers filed suit against him claiming that he had violated the spirit and intent of the Keystone agreement.

In July, only five months after Olney and Audenreid were slated for reconstitution, an independent arbitrator ruled in the favor of the PFT on four out of the five issues he was asked to consider. The arbitrator found that the Superintendent had not met the minimum threshold of "partnership" in working with the Philadelphia Federation of Teachers to help improve student achievement at Olney and Audenreid. He noted that the Superintendent failed to appoint a District liaison to regularly work with the PFT on district reorganization, failed to notify the PFT at "the earliest practical date" of the decision to reconstitute (calling the head of the PFT only the night before the announcement), failed to notify the PFT of meetings being held to determine the selection criteria for distressed schools, and failed to give the PFT its fair say in criteria for selecting one-quarter of each Keystoned school's faculty. As a result of this decision, the schools were not forced into reconstitution.

Despite the arbitrator's unfavorable ruling, Superintendent Hornbeck believes the Keystone controversy was beneficial in that it focused intense public attention on student achievement, attendance, graduation and promotion. He noted that few issues have penetrated District perceptions as much. As one principal explained, "The events surrounding the reconstitution of Olney and Audenreid were like a 'lightening bolt' because given the scores of those schools, it was clear, it could be us."

While the reconstitution announcement may have acted as a catalyst to other schools, it had equally strong but harder to assess effects in the two targeted schools. Protest against the reconstitution announcement came in many forms. Olney students staged walk-outs during school for three straight days. Olney also experienced a series of incidents including the theft of the principal's files relating to the reorganization process, vandalism, and the sealing of fire doors to cause disruption. The PFT building representatives at the two schools circulated a letter declaring that any teacher who reapplied to stay in the school was a "scab." And a petition was circulated calling for the removal of the principal.

To say that teacher and student morale at the reconstituted schools was low would be an understatement. As one teacher expressed, "This is a serious situation, the keystoning. It is affecting the teachers and the kids. The climate here is negative and pervasive. We are losing control of the students." Another teacher said, "I think this Keystone designation represents a betrayal on the part of the Superintendent. I don't think he has a clue or a plan."

The long-term impact of the Keystone designation of the two schools remains unclear. For the short term, it has been disruptive and diffusive for the two schools. Both schools have since undergone leadership changes, and many staff members have transferred out. Since ultimately neither Olney or Audenreid was reconstituted, time will tell if the temporary designation forced change in the schools. There may be a broader impact, however, on attitudes within the system about the seriousness of the district's commitment to reform and the intent to act when there is insufficient progress.

PRELIMINARY ASSESSMENT *of the* SYSTEM

While it is too early to tell whether the accountability system in Philadelphia will positively impact student achievement over the long run, it can be assessed in terms of research which outlines the features of an effective accountability system. The Philadelphia components can be examined using these criteria. One effort to synthesize current thinking on this issue was recently undertaken in Delaware. In 1996 the Business/Public Education Council of Delaware charged the Delaware Education Research & Development Center with summarizing the best national thinking on accountability. Numerous prominent education researchers and policymakers (including David Hornbeck) knowledgeable about accountability were interviewed and a set of principles was developed. These principles are listed below in Chart 3 along with some others derived from other research on accountability. (See Meyer, 1994.)

CHART 3

Principles of an Effective Accountability System

1. The accountability system should be easily understood by, and make good sense to, the public and educators.
2. All system participants—students, parents, educators, schools, business and the community—should be accountable in some way.
3. Performance should be linked to consequences for individuals as well as schools and districts. Schools whose students meet or exceed the standards should be rewarded. Those that need help should receive it. Those that persistently or dramatically fail should be penalized.
4. The system should motivate people to act to achieve the desired goals.
5. An accountability system must include help for students having difficulty meeting the requirements.
6. Accountability should be tied to progress toward the state's academic standards, with students and schools expected to perform at agreed-upon high levels.
7. Progress toward the standards should be measured with technically adequate and fair assessment tools.
8. The assessment should be comprehensive so as to ensure all children access to the full curriculum.
9. The assessments should reflect and encourage good instructional practice.

These principles are used below to assess the components of the accountability system in Philadelphia. This section of the paper discusses what has been done with regard to each principle, how well the system satisfies each of them in the eyes of various stakeholders, and further issues that need to be addressed.

1. The accountability system should be easily understood by, and make good sense to, the public and educators.

In our interviews with them, most cluster leaders and principals and many other educators in the School District of Philadelphia revealed at least a basic understanding of the PRI. They were aware that it was being used for accountability purposes, knew that their score was dependent on how many students participated in testing and how well students performed on the SAT-9 tests. Many were also cognizant that other factors played a role in the calculation of the PRI, with teacher attendance being the most often mentioned specifically. Most cluster leaders reported that they had met with their principals to discuss the PRI and to help develop plans to meet the initial target scores. Pointing out disgraceful attendance and testing statistics for one of her schools, a cluster leader told us, "[The principal] knows he has to move, and we [the cluster staff] have to help."

In our 1997 survey, slightly more than half of Philadelphia public school teachers reported that they understood the purpose of the PRI. Combining this statistic with data from our interviews, this lack of understanding seemed to be more of an indicator of skepticism about the purpose of the PRI, rather than a measure of true confusion. To be sure, it was clear in our interviews with teachers that they were aware of pressure to increase test scores and improve student attendance at testing. In one classroom in which students were practicing test-taking skills using the Key Links workbooks, the teacher clearly identified to the evaluation team member present which students would be tested. Noticing one girl who did not have a workbook, the researcher offered the workbook she was following along with to the girl. "She's not tested. ESOL, level 1," responded the teacher.

Although the public and various system participants seem to embrace the idea of accountability in Philadelphia's schools, it is not clear how well the actual PRI is understood by the stakeholders. While most people are aware of the elements of the index—SAT-9 achievement levels, student and teacher attendance, and promotion/graduation rates—most would probably not recognize the mathematical formulae which are used to compute the overall score. As discussed previously, there are layers of complicated calculations involved. Certainly most people (including the public and teachers) do not know why each component of the PRI is weighted as it is, or why each school needs to make equal increments of progress toward their targets every two years over a twelve-year period.

Additionally, not all the elements of the PRI are seen as fair by many stakeholders. For example, the staff attendance variable is an issue of some controversy. All staff that are assigned to a particular school and report to the principal are included in this variable. This includes teachers, the principal, librarians, reading teachers, nurses, paraprofessionals, noon-time assistants, school security, etc. All absences for illness, illness in the family, and personal leave are included in the calculation for staff attendance. What this means in practice is that a teacher out on maternity leave or with a prolonged illness will bring down his or her schools' attendance score. District officials argue that the staff attendance rate should not "represent a judgement about the legitimacy of the absence, but simply measure how many days staff are present to provide services to students." While this may seem logical, to a principal in a small school with two or three teachers out due to extended illness, it seems unfair.

43

2. All system participants—students, parents, educators, schools, business and the community—should be accountable in some way.

Hornbeck has stated repeatedly, "Accountability must not apply to educators alone. Parents, students and the wider citizenry also are responsible." The goal of the accountability system under *Children Achieving* is to make all stakeholders more accountable for student performance, but this has not yet been accomplished. Who is affected by the accountability system and how is outlined in Chart 4 below.

CHART 4

Effects of the Accountability System on Various Stakeholders

Schools	Can receive awards (up to $1,500 per teacher) if PRI scores improve; if scores continue to decline, schools can be reconstituted
Superintendent	Salary increase tied to whether or not systemwide achievement increases; superintendent can be penalized up to 5 percent of pay for declines
Cabinet	Salary increase tied to whether or not system wide achievement increases; Cabinet can be penalized up to 5 percent of pay for declines
Students	Increased promotion/graduation requirements in year 2000
Parents	May be required to sign a contract agreeing to ensure good attendance, provide health needs and encourage reading (No enforcement)
Citizens	Must produce significant increase in money (No enforcement)

As the above summary demonstrates, while there are plans to hold all stakeholders accountable, the requirements for students, parents, and the wider citizenry have not yet been implemented. The School District has pledged not to put these components of the responsibility system into effect unless new supports—money, volunteers, etc.—are put in place first. However, given the School District's budget difficulties, it is not clear when or if this will happen. As justification, District officials argue that:

Teachers, principals and other school staff are primarily responsible for student achievement, because they can have enormous influence on student learning and supporting student achievement is what their jobs are all about. As the School District works with parents and student groups to develop statements of their responsibilities, it is appropriate for us to affirm our own responsibilities first.

Even though District officials, teachers and principals are currently being held accountable, Hornbeck has said that if the School District continues to be underfunded, the entire accountability system will be suspended in the year 2000.

3. **Performance should be linked to consequences for individuals as well as schools and districts. Schools whose students meet or exceed the standards should be rewarded. Those that need help should receive it. Those that persistently or dramatically fail should be penalized.**

Consequences are clear for individuals and schools in Philadelphia. The Superintendent and the Cabinet members have salaries tied to student performance while teachers and principals risk forced transfer. Additionally, entire schools can receive extra funds or risk reconstitution depending on student achievement.

4. **The system should motivate people to act to achieve the desired goals.**

The *Children Achieving* theory of action of instructional change holds that if the District specifies high academic standards for students as a focus for the efforts of teachers and administrators and couples this with a high-stakes accountability system to provide incentives, with adequate support, teachers will do what is necessary to help students achieve those standards. The question then becomes: Does the accountability system provide sufficient motivation?

The rationale behind the accountability index is that it will work to improve schools in one of four ways. Either teachers will teach better because:

- they can receive an award (either cash or public recognition) for higher test scores;
- they are motivated to take advantage of the assistance being provided by the Teaching and Learning Network and other support structures;
- the PRI clarifies practices and promotes coordination; or
- they will be sanctioned in various ways if student performance fails to improve.

The implication is that if teachers work harder and smarter, student achievement will improve. It is simply too soon to tell whether these incentives will lead to sustained improvements in student achievement. In 1996-97 most of the system, 16 of the 22 school clusters, were in their first year of implementing the components of *Children Achieving*. The standards and rewards and sanctions associated with the Professional Responsibility System were new to the teaching staff and the incentives associated with the system had not had time to affect policies and practices in many of the schools. It is probably the case that the idea of rewards was simply an abstraction to many teachers in 1996-97 and that it will remain so until the rewards have been distributed for the first time. Therefore, judgements about the impact on teaching practice and student achievement will have to be addressed in the future when further data has been gathered and analyzed.

5. **An accountability system must include help for students having difficulty meeting the requirements.**

Before implementing the new promotion and graduation requirements, District officials want a support process in place in every school with summer school and extended days for failing students. If such a system is put in place, it certainly will help students that are having difficulty meeting the requirements. However, new standards and a new assessment are already in place, but the student support system is not. Teachers and administrators are being held accountable without the assistance of extra time for failing students.

It is the District position that the structural elements of the *Children Achieving* reform serve as adequate supports to help schools reach their targets at least until the year 2000. At that time, if no new resources are brought into the system, the entire accountability system will be suspended. Below is a description of the current support elements:

- **Standards:** The Philadelphia standards support basic skills, create a capacity for life-long learning, and reflect the heritage of the people in our city.
- **Full-day Kindergarten:** All elementary schools have a full-day kindergarten program for eligible children. This provides increased time for teachers to provide the necessary academic, social and motor development that will help young children succeed in the regular grades.
- **Clusters:** By creating an organization that allows for planning and decision making around the entire period of a child's education, a system has been created which will allow deep-rooted changes that will raise achievement for all children.
- **Teaching and Learning Network:** The TLN provides professional development activities and direct in-classroom support to insure implementation of standards-based activities and use of performance assessment tasks. In addition, it facilitates K-12 articulation among principals, teachers and parents.
- **Increased Professional Development:** The District has increased the amount of professional development available to each teacher, administrator and staff member.
- **Equity Support:** Each cluster has an Equity coordinator who works in collaboration with the cluster team to ensure that all students served by the District have equal access to educational opportunities for success in achieving rigorous standards.
- **Small Learning Communities:** These are schools-within-a-school where all staff and students share a clearly defined sense of purpose. They will be heterogeneous and committed to enabling all students to achieve rigorous standards and will be accountable for student outcomes and have decision-making authority equal to that responsibility.
- **Family Resource Network:** Provides a system of support in the areas of health, safety and attendance for students and their families.
- **More Books and Technology:** Increased funding and local decision making allow schools to select the books and materials that best meet the instructional needs of their school.

6. Accountability should be tied to progress toward the state's academic standards, with students and schools expected to perform at agreed-upon high levels.

The factor given the most weight in the accountability system is the Stanford-9 Achievement Test, and many teachers in the Philadelphia system do not perceive this assessment as adequately aligned to the new standards adopted in the city. On the survey, only 35 percent of teachers reported that they believe the SAT-9 accurately reflects the standards. In fact, the District's own curriculum experts found that the SAT-9 was not a good match with the standards in all subject areas. However, District officials argue that the test was "good enough for the broad baseline purposes for which we used it, and for the first cycle of the professional responsibility system."

The central administration in the District is trying to remedy part of the problem by developing resource guides. After the standards were released in 1996, the Office of Curriculum Support issued a set of Standards Curriculum Resources Guides for grades K-4, 5-8, and 9-12 in english/language arts, math and science. These documents were meant to replace the previous administration's standardized curriculum. However, they received mixed reviews in the schools both as to their usefulness and their actual alignment with the standards. As a result, the District just recently (January 1998) released the more explicit Curriculum Framework for all core subject areas. These guides are intended to help teachers make the connection between the standards and helping students achieve on the assessment instrument. How well these guides meet that goal will be explored in the spring 1998 evaluation fieldwork.

In addition, School District officials have been working with Harcourt Brace, the publisher of the SAT-9, to develop new test questions which may better align the assessment with the Philadelphia standards. Some of these items were included on a pilot basis in the 1997 administration of the test and are part of the test battery being administered in Spring 1998. However, whether or not these new items result in better alignment with the standards remains to be seen.

7. Progress toward the standards should be measured with technically adequate and fair assessment tools.

The Stanford-9 Achievement Test is certainly recognized in the educational testing industry as a technically adequate and valid measure of student performance. It is used in many large districts, including Boston, Los Angeles, and Houston. However, questions have been raised about its alignment with the standards (see above) and about the quality control procedures of the test publisher. In addition, more fundamental questions have been raised about how the scores are being used in Philadelphia.

In January 1998 the Superintendent announced that Harcourt Brace made a scoring error that led the District to slightly underestimate how well students were achieving. The error misrepresented the 1996 levels of achievement on the exam for just two percent of the students tested. However, the mistake was large enough that two schools which had been targeted as low progress (because their 1997 scores fell below their baseline scores) were taken off the distressed list when the test scores were corrected. Also, because the revised 1996 citywide scores are lower than originally calculated, overall improvement from 1996-97 is actually slightly better than originally reported.

In addition to questions about the assessment, other concerns have been expressed about the validity of using an index to rate performance. One implication of the PRI is that it is reasonable to expect schools to raise their test scores in all subjects 12 years in a row. For some schools reaching the district goal of 95% of students performing at or above the standards means very large gains in achievement. Another implication is that schools can sustain these rates of growth, cycle after cycle. There is no precedent for such gains and no empirical basis for calculating reasonable targets. Some scholars have questioned this approach, arguing that the targets are not attainable (Koretz, 1998).

Another issue cited by detractors is that different cohorts of students are tested each year and that a school's progress is measured by comparing the performance of different groups of students. The assumption is that one cohort of students looks similar enough to another that the salient variable in achievement is instruction. This assumption is problematic in especially small schools where small differences in cohorts are likely to result in rather dramatic differences in test scores.

8. The assessment should be comprehensive so as to ensure all children access to the full curriculum.

The SAT-9 is now administered to almost all students in grades 2, 3, 4, 7, 8, 10 and 11 in reading, math and science. The only students exempted are those who are classified as severely and profoundly impaired, as trainable mentally retarded, autistic or are in ESOL at Level 1. Any student who does not complete all three sections of the test is given a score of zero which affects how a school performs in terms of the PRI (discussed below). This is to ensure that school administrators do not "inflate" their scores by testing only those students who they believe will perform well. Because all students must take the exam, administrators in Philadelphia are attempting to ensure that all students will have access to the full curriculum.

9. The assessments should reflect and encourage good instructional practice.

As discussed previously, because the Stanford-9 Achievement test counts for the majority of the total score in the Performance Responsibility Index, teachers and schools have made an all out effort to increase student achievement on the test. Whether or not this qualifies as "good instructional practice" is a matter of some debate. As outlined in the findings section, since the SAT-9 counts for a large portion of the overall PRI score, school staff have made no secret of the energy they are directing to improving student test scores. Is this "teaching to the test" a desired outcome? Although many teachers believe that the test has encouraged them to develop their students' critical thinking and problem-solving skills, will this really improve the quality of pedagogy? Will it have a real impact on student learning over time? How much of the gains on the SAT-9 between 1996 and 1997 can be attributed to test effects? As one teacher said, "We have plans for teaching the skills necessary for various components of the test, but it has distracted our focus to once again using traditional means for traditional targets to satisfy the short-run goals."

Summary

The above discussion demonstrates that the School District does well on some of the accountability criteria, but has some work left to do before it can be judged a success. The complexity of the accountability system is not well understood, not all system participants are held equally accountable, more supports need to be available to help students achieve the standards, and how measures on the index are calculated needs to be re-thought. However, the School District has made consequences clear for all participants and the assessment system is deemed by most to be technically adequate and fair. The following recommendations are intended to assist District officials in improving the accountability system in general.

RECOMMENDATIONS

- An expert panel should be appointed to review the PRI, monitor it over time, and advise the District. This will help District officials prove the quality of the system and build public confidence in the results. This may also reassure teachers that they are not being treated unfairly.
- The School District should put more effort into explaining the components of the PRI. Although only 51 percent of teachers reported understanding the purpose of the PRI on the teacher survey, 63 percent of those teachers believed that it has the potential to benefit their students. This suggests that if District officials make a greater effort to educate teachers about the purpose of the PRI, more may see the potential benefit.
- The calculation of the staff attendance variable should be changed so that long-term illnesses are no longer included. The qualitative data suggests that this may improve teacher and administrator attitudes toward the PRI.
- The SAT-9 categories (Advanced, Proficient, Basic, and Below Basic) should not be used for the non-cognitive indicators in the accountability index because there is no empirical basis for the cut points used to assign schools to these categories. Their use eliminates much of the actual variation, while it exaggerates the importance of minor differences among schools. Long-term targets might be based on state or national data and interim targets could be set based on reasonable progress toward those goals. The actual changes in performance could be used in the index.
- A local panel of experts (including teachers and the PFT) should work with the SAT-9 test publisher, Harcourt Brace, to review the alignment of the revised SAT-9 with the Philadelphia standards.
- School District officials should move with due haste to pilot additional student performance indicators that can supplement the SAT-9, such as portfolios and course exams.
- New programs often take more than two years to produce effects or increase achievement. Thus, a school that has adopted an appropriate course of action and is working hard to implement it may fail to reach its numerical target. For this reason, when District officials publicly identify schools as "low progress," they should include information about action taken to improve performance.

REFERENCES

Abelman, C., Elmore, R. & Kenyon, S. (1997). Local meanings of accountability. Paper presented at the annual meeting of the American Educational Research Association.

Anyon, J. (1997). *Ghetto schooling: A political economy of urban education reform.* New York: Teacher's College Press.

Fine, P., LeMahieu, P. & Perry, C. (March 1997). Building Effective Accountability and Professional Development Systems for Delaware's Public Schools: Phase One Report. Prepared for the Delaware Business/Public Education Council.

Ladd, H. (1996). Introduction. In H. Ladd (Ed.), Holding schools accountable: Performance-based reform in education. Washington, DC: The Brookings Institute.

Hedges, L., Laine, R. and Greewald, R. (1994) Does money matter? A Meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(4), 5-14.

Meyer, R. (1996). Comments on chapters two, three, and four. In H. Ladd (Ed.), Holding schools accountable: Performance-based reform in education. Washington, DC: The Brookings Institute.

ABOUT *the* CHILDREN ACHIEVING CHALLENGE

Many innovative school reform plans have foundered for lack of resources. In February 1995, shortly after the School Board adopted Children Achieving, The Annenberg Foundation designated Philadelphia as one of a small number of American cities to receive a five-year, $50 million Annenberg Challenge grant to improve public education.

Among the conditions for receiving the grant was a requirement to produce two matching dollars (i.e., $100 million over five years) for each one received from the Annenberg Foundation, and to create an independent management structure to provide program, fiscal and evaluation oversight of the grant. To assist in meeting both these conditions, the District turned to Greater Philadelphia First, an association of chief executives from the region's largest companies, to help raise the matching dollars and to provide the oversight required by The Annenberg Foundation. A staff was hired, and the Children Achieving Challenge came into being.

For the Challenge staff, the initial question was how to harness the, at times, fragmented efforts of various organizations that work with the School District to improve schools. Such organizations usually focus on specific projects but often have been unable to do much to improve the school system as a whole. For this reason, Challenge staff have served as catalysts, conveners and coordinators in a massive collaboration between internal and external partners. As a result, the Challenge has helped bring the School District together with all of its potential partners in a collective focus and a new way of working that can sustain itself long after the Challenge is gone.

Greater Philadelphia First houses the Challenge and provides oversight to it through the GPF Partnership for Reform. In addition to its focus on education, GPF provides leadership on issues important to the economic development and quality of life of the community.

Children Achieving Challenge
c/o Greater Philadelphia First
1818 Market Street, Suite 3510
Philadelphia, PA 19103-3681
Phone 215 575 2200
Fax 215 575 2222

OVERSIGHT COMMITTEE *of the* CHILDREN ACHIEVING CHALLENGE

Thomas F. Donovan, Committee Chair
Chairman & CEO, retired, Mellon PSFS

Susan B. Butler
Managing Partner, Andersen Consulting LLP

Dr. Frederick W. Capshaw
President, Community College of Philadelphia

Yvonne M. Epps
President, Philadelphia Home and School Council

Alan R. Hirsig
President & CEO, ARCO Chemical Company

Ronald J. Kozich
Area Managing Partner, Ernst & Young LLP

Wayne O. Leevy
Managing Partner, Mitchell, Titus & Company

The Reverend Issac J. Miller
Church of the Advocate

Emanuel Ortiz
Deputy Mayor, City of Philadelphia

Dr. Judith Rodin
President, University of Pennsylvania

Richard L. Smoot
President & CEO, PNC Bank

Roger J. Steinbecker
Managing Partner, Price Waterhouse LLP

A WORD *of* THANKS

The Children Achieving Challenge would like to acknowledge the financial support of Philadelphia-based and national corporations and foundations in the funding of this evaluation. In particular, the technical assistance and funding support provided by The Pew Charitable Trusts has been extremely helpful.